Distributed video coding and its application to error resilient video communication

Zhang, Yixuan

2013

Zhang, Y. (2013). Distributed video coding and its application to error resilient video communication. Doctoral thesis, Nanyang Technological University, Singapore.

<https://hdl.handle.net/10356/54842>

<https://doi.org/10.32657/10356/54842>

DISTRIBUTED VIDEO CODING
AND ITS APPLICATION
TO ERROR RESILIENT
VIDEO COMMUNICATION

ZHANG YIXUAN

School of Electrical and Electronic Engineering

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

2013

Acknowledgements

First of all, I would like to thank my supervisor, Prof. Zhu Ce, for his inspiration, encouragement and patience during my candidature. The academic rigor and creative thinking learnt from him will benefit my entire life. My thesis cannot be completed without his encouragement and guidance. Apart from being an excellent supervisor, he is also a sincere and generous friend.

Next, I would like to acknowledge School of Electrical and Electronics Engineering, Nanyang Technological University (NTU) for providing me financial support. Years of studying and living in the garden-like campus are my best memories forever.

Furthermore, I want to take this opportunity to express my deepest gratitude to those who have provided guidance during these years. I would like to thank Prof. Yap Kim Hui for his constructive comments. I also would like to thank Prof. Zeng Wenjun, Prof. David S. Taubman and Prof. Xiong Zixiang for their professional insight and technical advices.

Throughout my Ph.D studies, friendships play an important role. I am grateful to Dr. Liu Minglei, Dr. Xu Zhiming and Dr. Wei Shikui for their help and shared experience. Thanks to Dr. Xu Yuanyuan, Dr. Xia Xu and Dr. Fan Xiaojia who share both the happiness and sorrow with me.

Finally, I would like to dedicate my deepest appreciation to my beloved wife and dearest parents who understand, support and believe in me throughout the years.

Contents

Contents	ii
List of Figures	vi
List of Tables	viii
List of Abbreviations	ix
Abstract	x
1 Introduction	1
1.1 Background	1
1.2 Motivations	2
1.3 Major Contributions	5
1.4 Organization of the Thesis	7
2 Literature Review	9
2.1 Fundamentals of Distributed Source Coding	9
2.1.1 Slepian-Wolf Theorem for Lossless Distributed Source Coding	9
2.1.2 Wyner-Ziv Theorem for Lossy Distributed Source Coding .	11
2.2 Distributed Video Coding	12
2.2.1 Background	12
2.2.2 Advance in Recent Years	13
2.3 Anatomy of Distributed Video Coding	15
2.3.1 Practical Distributed Source Coding	15
2.3.2 Side-Information Estimation	18

2.3.3	Source Correlation Estimation	21
2.3.4	Quantizer Design	23
2.3.5	Transmission Rate Control	24
2.4	Emerging Applications of Distributed Video Coding	25
2.4.1	Error Resilient Video Coding	25
2.4.2	Flexible Video Decoding	26
2.4.3	Parallel Video Encoding	26
3	Adaptive Nested Scalar Quantizer for Distributed Video Coding	28
3.1	Introduction	28
3.2	Fundamentals of One-Dimensional Nested Quantization	30
3.2.1	Theoretical Results on 1-D Nested Lattice Quantization . .	32
3.2.2	$Mod()$ Coset Partition	33
3.2.3	Scalar Quantizer (SQ) Based Coset Partition	34
3.2.4	1-D Nested Quantization for Video Coding	35
3.3	Distributed Video Coding Using Adaptive Nested Scalar Quanti- zation Based on Empirical Analysis	36
3.3.1	Motivation	36
3.3.2	Encoder	39
3.3.2.1	Discrete Cosine Transform and Quantization . . .	39
3.3.2.2	Mode Classification For Bit Allocation	39
3.3.2.3	Encoding With The Adaptive Coset Partition Method	41
3.3.3	Decoder	42
3.3.3.1	Identifying Possible Values	42
3.3.3.2	Finding Side-Information Block	43
3.3.3.3	Reconstructing Wyner-Ziv Block	43
3.3.4	Experimental Results	43
3.4	Distributed Video Coding Using Adaptive Nested Lattice Quan- tization and Code Design Based on Laplacian Correlation Noise Modeling	47
3.4.1	Limitations of Nested Lattice Quantization Based on Lapla- cian Correlation Noise Modeling for Video	47
3.4.2	Proposed Adaptive Nested Quantization	49

3.4.3	Code Design	52
3.4.4	Implementation of the Adaptive Nested Quantization Scheme	55
3.4.5	Experimental Results	59
3.5	Summary	62
4	Exploiting High-Order Correlation in Distributed Video Coding	65
4.1	Introduction	65
4.2	The Distributed Video Coding Framework	67
4.2.1	Decoder Multi-Resolution Motion Refinement (MRMR) Algorithm	68
4.2.2	The Laplacian Distribution Based Correlation Estimation .	70
4.3	Proposed Correlation Estimation Scheme	71
4.3.1	Motivation of high-order correlation exploitation scheme .	72
4.3.2	The Proposed Inter-coefficient Correlation Modeling	73
4.3.3	The Proposed Inter-bit Correlation Based Prediction Process	78
4.3.4	The proposed inter-bit correlation based prediction process	78
4.3.5	Integrated Symbol A Priori Probability Prediction	80
4.4	Experimental Results	81
4.5	Summary	84
5	A Joint Source-Channel Video Coding Scheme Based on Distributed Source Coding	87
5.1	Introduction	87
5.2	Review of DSC-based Error Resilient Video Coding Schemes . . .	90
5.2.1	DSC-based Error Resilient Predictive Video Coding	90
5.2.2	DSC-based Error Resilient Fine Granularity Scalability Video Coding	91
5.3	Analysis of the DSC-based Joint Source-Channel Coding Scheme .	92
5.3.1	Efficiency Analysis of DSC-based Joint Source-Channel Coding vs. Predictive Coding with FEC in the Case of Channel Errors at the Reference Only	92
5.3.2	Comparative Efficiency Analysis in a More General Case .	96
5.4	DSC-based Joint Source Channel Coding Design for Video	98

5.4.1	Encoder	100
5.4.1.1	Motion Estimation and Prediction	100
5.4.1.2	DCT and Quantization	101
5.4.1.3	Forming Bit-Planes	101
5.4.1.4	Joint Source-Channel Encoding with LDPCA Code	101
5.4.2	Decoder	102
5.5	Simulation Results	103
5.6	Summary	106
6	Conclusions and Future Works	109
6.1	Conclusions	109
6.2	Recommendations for Future Work	111
6.2.1	Side-information Estimation	111
6.2.2	Correlation Estimation	112
6.2.3	Rate-distortion Optimization for DVC	112
6.2.4	Error Resilient Video Coding	112
6.2.5	Removal of feedback channels in SW coding	113
	Author's Publications	114
	Bibliography	116

List of Figures

2.1	Comparison between centralized source coding (CSC) and distributed source coding (DSC).	10
2.2	Practical Wyner-Ziv coder.	12
2.3	Low complex encoding and decoding with transcoder.	13
2.4	A block diagram for the DVC system.	16
2.5	An example of DSC using binning.	16
2.6	Motion compensated extrapolation and interpolation at the decoder side.	19
2.7	DVC for flexible multi-view video decoding.	26
2.8	Serial encoding using CVC vs. parallel encoding using DVC. . . .	27
3.1	An illustration of the 1-D nested lattice quantization.	31
3.2	Practical 1-D nested lattice quantization scheme.	31
3.3	An example of $Mod()$ coset partition with $N = 3$ and $M = 18$. . .	34
3.4	An example of SQ-based coset partition with $q = 6$	35
3.5	The relationship between averaged co-difference and the correlation noise at the block level for four different sequences.	38
3.6	Proposed distributed video coding scheme.	39
3.7	Rate-distortion comparisons (H.264 baseline codec JM12.4 is used with RDO off).	44
3.8	Frame based PSNR comparison of different coset partition methods. .	45
3.9	Visual quality comparison of different coset partition methods (foreman.cif, frame 180).	46
3.10	Approximation of correlation noise Z	50
3.11	Proposed adaptive 1-D nested quantization.	51

3.12	Distribution of block-wise correlation noise $ Z $ for “foreman.qcif” sequence.	52
3.13	Rate R in (3.14) as a function of T with varying q ’s and α ’s. . . .	56
3.14	Plot of optimal T and N as a function of α with varying q ’s. . . .	57
3.15	Correlation noise distribution with different GOP sizes of 2, 4 and 8. 60	
3.16	Rate-distortion performance comparison of the proposed scheme against conventional DVC scheme with different GOP sizes of 2, 4 and 8.	64
4.1	Decoder multi-resolution motion refinement (MRMR) algorithm. .	68
4.2	The distribution of the coefficients given their parent magnitudes.	75
4.2	The distribution of the coefficients given their parent magnitudes.	76
4.2	The distribution of the coefficients given their parent magnitudes.	77
4.3	Context window.	79
4.4	Rate-distortion performance comparisons for different sequences.	85
4.4	Rate-distortion performance comparisons for different sequences.	86
5.1	The error resilient video compression framework in [1].	88
5.2	DSC-based error resilient FGS video coding in [2].	91
5.3	An example of error resilient compression using coset partition. .	94
5.4	Comparisons of $R_{[1]}^{peg}$ and R_{joint}^{peg}	99
5.5	The proposed encoder for DSC-based joint source-channel video coding.	100
5.6	The proposed decoder for the peg frames.	102
5.7	Decoded PSNR comparisons with $QP = 28$ and real BSC crossover probability $p_r = 0.01$	104
5.8	Peg frame bit rate comparisons with $QP = 28$ and $p_r = 0.01$. . .	105
5.9	Peg frame rate-distortion comparisons with different schemes ($p_r = 0.01$).	107
5.9	Peg frame rate-distortion comparisons with different schemes ($p_r = 0.01$).	108

List of Tables

2.1	Ratio of encoding time between H.264 and DVC system in [3] (reported in [4])	14
3.1	Level 1 - bit allocation per Wyner-Ziv block.	41
3.2	Level 2 - bit allocation in percentage to the two coset partition methods.	41
3.3	Level 3 - bit allocation in percentage for each DCT coefficient $C(i, j)$ per block.	42
3.4	Statistics comparisons of the actual correlation noise Z against the Laplacian approximation with $\alpha = 0.35$ for “foreman.qcif” sequence.	49
3.5	Selection of N with respect to q for different sequences	61
4.1	KLD for different distributions approximation	78
4.2	Prediction results and coding rates with example of the first fif- teen coefficients in the frequency band HL_2 of the third frame of “foreman.qcif”	82

List of Abbreviations

Abbreviation	Description
1-D	one-dimensional
BER	bit-error-rate
CSC	centralized source coding
CVC	centralized video coding
DSC	distributed source coding
DVC	distributed video coding
FEC	forward error correction
GOP	group of picture
i.i.d	independently identically distributed
JSCC	joint source-channel coding
KLD	Kullback-Leibler divergence
LDPC	low-density-parity-check
LDPCA	low-density-parity-check-accumulated
LSB	least significant bit
MMSE	minimum mean square error
MRMR	multi-resolution motion refinement
MSB	most significant bit
NSQ	nested scalar quantization
RD	rate-distortion
SI	side-information
SQ	scalar quantizer
SW	Slepian-Wolf
WZ	Wyner-Ziv

Abstract

In current centralized video coding (CVC) systems including all the ITU-T and MPEG standards like H.264/AVC, the encoder is designed to compress input signal by exploiting all the temporal, spatial and statistical redundancies presented in a video sequence. As a result, all the centralized video encoders are much more complex than their corresponding decoders. Such framework is suitable for broadcasting applications where compressed bitstreams are repeatedly served to users. On the other hand, according to Slepian-Wolf theorem and Wyner-Ziv theorem for lossless distributed source coding (DSC) and lossy DSC respectively, it is also possible to accomplish efficient signal compression by exploiting source correlation at the decoder only. By applying the DSC mechanism in video compression known as distributed video coding (DVC), we can shift the correlation exploitation tasks of high computational complexity from the encoder to decoder, especially that of exploiting temporal correlation in video coding, thus achieving low complexity encoding. This feature is highly desired in the emerging wireless video communication applications where compression needs to be accomplished at the encoders with the scarcity of computation and especially power resources, such as wireless video surveillance cameras, tablets and smart phone cameras, and remote video sensors.

Despite the great efforts spent on the research of DVC in the last decade, the rate-distortion performance gap between DVC and CVC remains significantly large. There are three main sources for the performance gap: less efficient channel-code based distributed source coding vs. entropy source coding, less reliable side-information estimation and less accurate correlation prediction. To enhance the performance of DVC systems in terms of coding efficiency, in this thesis, we consider quantizer design with side-information and high-order correla-

tion estimation. We first address quantizer design by studying two coset partition based quantization schemes for better reconstruction quality and lower bit rate, respectively. Meanwhile, by exploring the high-order correlation in wavelet domain to improve the correlation model accuracy, we lower the bit rate in DVC. More details are elaborated in the following.

One of the practical quantizer designs in DVC is the coset partition based one-dimensional nested lattice quantizer which use a scalar quantizer followed by a coset channel code. Two typical coset partition methods, namely, modular based coset partition and scalar quantizer based coset partition, are examined in this thesis. The modular based approach introduces no reconstruction error when the correlation noise is small, but suffers a large error when the correlation noise goes beyond the error correction capability of the coset channel code. In contrast, the scalar quantization (SQ) based binning has an advantage that the maximum decoding error can be clipped in a certain range even though there is an unexpected large correlation noise, but errors may be inevitable even the correlation noise is small. Taking advantage of their respective strengths while circumventing their weaknesses, an adaptive coset partition scheme is proposed by integrating the two coset partition methods to minimize the decoding errors given a rate constraint. Different modes are designed empirically for bit allocation between the two coset partition schemes based on the correlation noise prediction. To further extend the study of quantizer design using coset partition to improve the performance, the nested quantization is investigated in a more analytical way. The adaptive nested lattice quantization scheme for distributed source coding is modified to minimize the rate for a given distortion. An analysis is presented to determine a correlation noise threshold based on which the two above mentioned coset partition schemes are employed adaptively. More specifically, we derive the overall rate of indices generated by different schemes with respect to the threshold, and then solve the problem numerically to find the threshold which minimizes the rate.

The correlation modeling is another key to the performance of DVC. In the most widely used channel code (low-density-parity-check (LDPC) codes and turbo codes) based DVC, correlation modeling is used to initialize the belief propagation decoding process by providing a priori probability estimates for the bits received.

A better estimation accuracy of the probability tends to improve the coding efficiency. As the existing Laplacian correlation noise modeling fails to exploit high-order statistical correlation, an effective approach is proposed to explore the inter-coefficient correlation across scales and inter-bit correlation within each frequency band in wavelet domain. Using the two levels of correlation exploitation plus the widely used Laplacian correlation noise modeling, we can achieve better a priori probability prediction through Bayesian approach. The proposed scheme is implemented in a recently developed wavelet domain DVC framework with significant and consistent coding gain achieved.

Apart from enhancing the coding efficiency of the DVC systems, we also investigate the error resilient video coding based on DSC. Error propagation is a common problem in conventional video coding in case a reference symbol is corrupted. It is known in DSC, the source coding can be considered as a virtual channel coding, where the reference is a corrupted version of the current source. Therefore a unified single channel code can be employed for joint source-channel coding which combats the virtual and real channel errors simultaneously. In this way, we propose a channel-aware error resilient video compression scheme using the DSC technique to stop error propagation. We apply a single distributed source code to certain frames for both compression and data protection purposes, thus stopping error propagation more efficiently. Compared with the separate source and channel coding, the proposed joint source-channel video coding scheme shows the rate saving in eliminating error propagation.

Chapter 1

Introduction

1.1 Background

In conventional centralized coding systems, encoders exploit correlation among source signals to achieve data compression. Consequently, all the centralized video encoders are much more complex than their corresponding decoders. Such framework is desired for broadcasting scenarios where compressed bitstreams are repeatedly served to users. However, a reversed architecture may also be required in some situations, where there are tight constraints to the encoders. For example, in contrast to server computers with constant power supply, wireless mobile devices usually operate using un-renewable energy throughout their whole missions. The availability of energy significantly impacts on such wireless systems, ranging from their data process capability, operation duration and distance for wireless communication, to the compactness of the devices. In such scenarios, the encoders with low complexity and power consumption are desired. Thanks to Slepian and Wolf's theoretic bound [5] for lossless distributed source coding (DSC), as well as Wyner and Ziv's bound [6] for lossy DSC, surprising but promising results were discovered, that is, high efficient compression is achievable by exploiting source correlation at the decoder only. These theorems lead to a new data compression architecture featured by separated (or distributed) encoders known as distributed source coding (DSC) [4]. By applying the DSC concepts to video compression, the correlation exploitation tasks can be shifted

from the encoder to decoder, especially for temporal correlation in video coding, thus achieving low complexity video encoding. This advantage is highly desired for these emerging wireless video communication systems where the encoders are limited by computational and power resources, such as wireless video surveillance cameras, tablets and smart phone cameras, and remote video sensors. Originally targeting at compact and long-life video encoders, distributed video coding (DVC) has become an interesting research paradigm based on distributed source coding.

Motivated by low encoding complexity, the distributed video coding was initiated in the beginning of this century, and is attracting increasing research interests with many representative DVC schemes in the past decade. However, the methodologies and techniques on DSC/DVC are not yet mature with some daunting technical challenges. The state-of-the-art DVC algorithms still cannot provide satisfactory or desirable solutions compared with conventional centralized video coding (CVC) schemes. That is, there exists a significant rate-distortion (RD) performance loss of the current DVC systems compared with the CVC standards like H.264/AVC [7] with low encoding complexity constraint. Although there is still a performance gap, DSC has recently find its new application to error resilient video coding, where better performance is reported in the presence of transmission channel error.

This thesis aims to investigate the DVC techniques and enhance the performance of the DVC systems for different contexts. By taking a divide-and-conquer strategy to dissect a typical DVC system and delve into each individual constituent, effective approaches are proposed to some of the constituents to improve the coding efficiency. Furthermore, the desirable feature of error resilience due to the unique DVC mechanism is explored to enhance video coding and streaming performance in the event of channel errors. In the following, the motivations and major contributions of this thesis will be elaborated.

1.2 Motivations

Compared to the centralized video coding schemes, there is still a significant performance loss when applying distributed video coding in terms of rate-distortion. This performance gap can be explained by three factors, namely, less efficient

channel-code based distributed source coding vs. entropy source coding, less reliable side-information estimate and less accurate correlation prediction. The channel codes based distributed source coding inefficiency comes from the inefficiency of the channel codes used in the DSC system. A performance loss is introduced if the channel code used cannot achieve the Shannon limit [8]. The side information and correlation model inaccuracy is due to the lack of necessary information (the symbols to be decoded) at the decoder (but before decoding) for accurate estimations. In this thesis, we focus on the study of the quantizer designs with side-information and the correlation modeling.

Nested lattice quantization provides a practical scheme for Wyner-Ziv quantizer design with side-information which is based on the coset channel code. The theoretical analysis of the nested lattice quantization based on coset channel code [9–11] has shown its potential to be the solution, but still assuming the high bit rate (high resolution) and a precise distribution model of the sources, which ensure that a “good” coset channel code can be constructed with a very small probability. However, in practice, the assumption of high rate may neither hold nor be used as a good approximation. Moreover, there is no accurate statistics model of the correlation between the source and its side-information in most practical scenarios like in distributed video coding. In these practical situations, such a “good” coset channel code is normally not available, leading to a much higher probability that a wrong code value would be selected to be the decoded result in a given bin. Thereby, the distortion with MSE metric becomes very large when a wrong code value is selected when decoding, since the codewords in a bin are far apart from each other. However, we note that the binning method based on the scalar quantizer (SQ) groups the neighboring code values into the same quantization bin, which restricts the range of code values in the bin. In this way, the maximum decoding error given the side-information can be clipped within a certain range even if the correlation between the source and its side-information is lower than expected. It can be seen that the selection of the conventional nested lattice quantization and SQ based binning should be made based on the correlation between the source and its side-information to maximize the coding efficiency and minimize the distortion.

The coding efficiency of the distributed video coding (DVC) systems highly

relies on the accuracy of the statistical correlation information. In contrast to conventional source and video coding, in distributed source and video coding, the decoder has no or inaccurate knowledge about the source correlation before decoding. Currently, the statistical correlation information is mainly obtained by the modeling of the correlation noise between a source and its estimated side-information [12–14]. Compared with the CVC which has employed entropy approaching tools to exploit high order statistical correlation, the only component to exploit the statistical correlation in DVC is the Laplacian correlation noise modeling at the decoder side. Such correlation exploiting scheme can hardly exploit high order statistical correlation. Therefore, approaches to exploit high order statistical correlation among the pixels, coefficients and their binary representations are highly desired in order to increase the coding efficiency.

Apart from enabling the low complex video encoding, DVC also exhibits error resilience in video coding due to its intrinsic coding mechanism. Conventional video coding with predictive coding is quite fragile to channel errors including packet erasure or bit errors. Such errors can cause predictive mismatch between reference used in the encoder and decoder. Drift effect is introduced as the mismatch goes on, which may result in severe or even disastrous quality degradation in the video decoding and reconstruction. In contrast, in DVC, since the predicted symbol is encoded independently on its reference symbol, robustness is naturally acquired against the reference errors. A few robust video coding schemes using DSC techniques have been proposed to prevent error propagation in predictive video coding [1, 15] and to protect enhancement layers in fine granularity scalability (FGS) coding [2, 16]. However, in these existing error resilient video coding schemes, the SW codes are only applied after predictive compression. Therefore, the SW codes are mainly or totally employed as channel codes, but ignoring their source coding capability. In view that the source coding can be done by considering the reference symbol as a corrupted version of the predicted (current) symbol through a virtual channel, while the source coding result has to be protected by channel codes when being transmitted through a real error-prone channel, it is therefore logical to consider fusing the virtual and real channels together to form a combined channel. Consequently, a single channel code can be applied to the combined channel to correct both the virtual and real channel errors, and better

coding efficiency can be expected.

1.3 Major Contributions

Motivated by the above issues, we have studied the coset channel codes, the nested lattice quantizer, and the high order correlation exploitation for enhancing distributed video coding efficiency, as well as the error resilient video coding based on DSC. The following contributions have been achieved.

In Chapter 3, we first examine two typical coset partition methods, namely, $Mod()$ coset partition (similar to the NSQ except the quantization) and scalar quantizer (SQ) based coset partition, where their respective advantages and limitations are investigated as described earlier. In order to maintain their advantages while circumventing their respective limitations, we propose an adaptive binning scheme by integrating the two coset partition methods. A distributed video coding system based on the proposed adaptive coset partition method is presented. The proposed coset partition method is a deterministic algorithm that adaptively allocates bits in the transform block coding based on a prediction of correlation noise for minimizing the decoding errors. Compared with other existing coset partition schemes as in [2, 10, 17] based on the $Mod()$ (NSQ) method, our proposed scheme can achieve better performance by integrating the two binning methods with an adaptive bit allocation.

In addition, we further study the nested lattice quantization by providing more analysis. We propose an adaptive 1-D nested lattice quantization scheme to employ the two above mentioned binning methods adaptively according to the correlation, for maintaining their respective advantages while mitigating their limitations. A correlation threshold is introduced in the selection of the two binning methods to minimize the rate for a given distortion. We also introduce an analytical way to determine the threshold based on a coarse modeling of correlation noise which can be obtained in practical video coding. More specifically, we derive the overall rate of indices generated by different binning schemes with respect to the threshold. Due to the difficulty in obtaining a closed-loop expression for the threshold which minimizes the rate, we solve the problem in a numerical way. We apply the scheme to distributed video coding and show the improvement

compared with the conventional DVC scheme.

In Chapter 4, wavelet domain DVC is considered, where we propose to predict the a priori probability by exploiting the inter-coefficient correlation across scales and inter-bit correlation within each frequency band, thus constituting a high-order correlation exploitation scheme to provide better accuracy to the a priori probability prediction. Specifically, with the exploitation of the self-similarity across scales, we can make better prediction of the coefficients of low magnitudes (insignificant coefficients), by modeling the correlation among the children coefficients and their parents. On the other hand, we also exploit the correlation among the current bit and the previously decoded neighboring bits in less significant bit-planes, which can help to predict the insignificance of the current bit especially in the higher frequency bands of smooth regions. With the two levels of correlation plus the widely used Laplacian correlation noise modeling, we can achieve an integrated symbol a priori probability prediction through Bayesian approach, thus yielding a better distributed source coding efficiency. The proposed scheme is implemented in a recently developed wavelet domain DVC framework, with decoder multi-resolution motion refinement (MRMR). Experimental results validate the effectiveness of the proposed scheme with significant and consistent coding gain over the original DVC system with MRMR.

In Chapter 5, a channel-aware joint source-channel video coding scheme based on DSC is proposed to eliminate the error propagation problem in predictive video coding in a more efficient way. It is known that near Slepian-Wolf bound DSC is achieved using powerful channel codes, assuming the source and its reference (also known as side-information) are connected by a virtual error-prone channel. In the proposed scheme, the virtual and real error-prone channels are fused so that a unified single channel code is applied to encode the current frame thus accomplishing a joint source-channel coding. Our analysis of the rate efficiency in recovering error propagation shows that the joint scheme can achieve a lower rate compared with performing source and channel coding separately. Simulation results show that the number of bits used for recovery from error propagation can be reduced by up to 10% using the proposed scheme compared with the existing DSC-based error resilient scheme.

1.4 Organization of the Thesis

The remainder of this thesis is organized as follows:

Chapter 2 provides a brief description of the background knowledge of the related topics by reviewing the relative literature. The review will cover the related topics discussed in the thesis, including the fundamental of distributed source coding, the practical distributed video coding, and emerging applications for distributed video coding.

Chapter 3 presents two adaptive coset partition schemes for distributed video coding. We start from discussing the two coset partition methods with their respective advantages and limitations. In the first work, we propose the adaptive integrated coset partition method to take their advantages while mitigating the limitations. The distributed video coding system using the proposed coset partition method is presented focusing on the bit allocation design. In the second work, we further investigate the adaptive scheme with more analysis based on Laplacian correlation noise modeling. Then we present our proposed adaptive scheme with the developed approach to determine the threshold of the proposed adaptive scheme based on a coarse correlation model for video coding. Experimental results comparing our schemes with related schemes are then provided to demonstrate the efficiency of our schemes.

Chapter 4 further investigates the coset partition in adaptive one-dimensional nested lattice quantization for distributed video coding. In this chapter, we first review the theoretical analysis and applications of nested lattice quantization. Then we present our proposed adaptive scheme with the developed approach to determine the threshold of the proposed adaptive scheme based on a coarse correlation model for video coding. Experimental results comparing with related schemes are then demonstrated to validate the efficiency of the proposed scheme.

Chapter 4 describes the developed approach to exploit high-order correlation in wavelet domain distributed video coding. We first review the related works of the latest decoder motion estimation technique. After that, we discuss the limitations of the existing Laplacian based correlation modeling, and then introduce the proposed high-order correlation estimation scheme, which considers the inter coefficient correlation across scales and the inter-bit correlation within each scale.

The experimental results using the proposed correlation exploitation method in distributed video coding are shown in the end.

Chapter 5 presents the proposed joint source-channel video coding scheme based on distributed source coding. We review and analyze some existing DSC-based error resilient video coding schemes. Then we provide analysis of the proposed DSC-based joint source-channel coding scheme. Next, the details of coding design for joint source-channel video coding are presented. Experimental results are shown to validate our scheme against the existing schemes.

Chapter 6 concludes the thesis and discusses the future works.

Chapter 2

Literature Review

In this chapter, we will review the fundamentals and various practical algorithms for distributed source coding (DSC), as well as the emerging applications based on distributed video coding (DVC) techniques.

2.1 Fundamentals of Distributed Source Coding

The history of distributed source coding (DSC) can be traced back to the 1970s, when Slepian and Wolf's pioneering work established the fundamental framework of DSC and introduced its theoretical performance bound. The DSC framework considers the encoding of two or more correlated but separated (or distributed) sources without communications among each other but decoding them jointly (shown in Fig. 2.1 as compared with conventional centralized source coding (CSC)). While each encoder compresses and transmits the coded bitstream separately (or distributedly), it is the decoder's job to exploit the statistical dependencies among the sources by jointly decoding all the received bitstreams.

2.1.1 Slepian-Wolf Theorem for Lossless Distributed Source Coding

With the reversed settings of conventional CSC, naturally, question arises: how is the performance of the new compression paradigm. The answer was provided in the theory established by Slepian and Wolf [5] that for lossless compression,

Figure 2.1: Comparison between centralized source coding (CSC) and distributed source coding (DSC).

DSC can achieve the same coding rate bound as CSC with an arbitrarily small error probability, provided joint decoding with the source correlation known to both the encoder and the decoder.

Mathematically, consider two statistically dependent finite-alphabet random sequences X and Y to be losslessly coded. Using the concept of random binning (also known as coset partition, and we use these two term interchangeably in this thesis) approach, Slepian and Wolf established theory [5] to show that the compression rate bound given in 2.1 can be approached with an arbitrarily small error probability using the DSC settings, which is the same as the well known Shannon's theory [18] about the minimum lossless coding rates for X and Y .

$$\begin{aligned}
 R_X &\geq H(X|Y) \\
 R_Y &\geq H(Y|X) \\
 R_X + R_Y &\geq H(X, Y)
 \end{aligned} \tag{2.1}$$

where $H(X|Y)$ and $H(Y|X)$ are the conditional entropies, and $H(X, Y)$ is the joint entropy of the sources.

In practical distributed source coding scenarios, especially in distributed video

coding, a special case known as the asymmetrical distributed source coding is of particular interest, where one source Y is coded independently and made available at the decoder. With the aid of Y which is known as side information (SI), the focus is therefore on coding the other source X with a rate as close to the bound $H(X|Y)$ as possible.

2.1.2 Wyner-Ziv Theorem for Lossy Distributed Source Coding

In 1976, Wyner and Ziv extended the Slepian-Wolf problem from lossless case to lossy case and studied the rate bound in asymmetrical DSC where the side-information Y is coded independently and available at the decoder only [6]. The rate-distortion (RD) function for DSC $R_{X|Y}^{WZ}(D)$ was investigated and it was found that for a given target distortion D , it holds in general that $R_{X|Y}(D) \leq R_{X|Y}^{WZ}(D) \leq R_X(D)$, where $R_{X|Y}(D)$ is the rate required to code X using CSC with Y known to both the encoder and the decoder, and R_X is the rate for independent coding of X (without side-information in neither the encoder nor the decoder). Wyner and Ziv have also shown $R_{X|Y}(D) = R_{X|Y}^{WZ}(D)$ for the zero mean, stationary, and memoryless Gaussian sources, with mean-squared error (MSE) distortion metric [6, 19]. Recent study showed that to achieve the identical rate-distortion performance as CSC, the DSC only requires the correlation between X and Y being Gaussian, which means X and Y can follow any arbitrary distribution. Meanwhile, Costas “dirty paper” theorem is the dual of DSC, which considers channel coding with the side-information available at the encoder only [20, 21].

Note that in many practical DSC scenarios like distributed video coding (DVC) which will be discussed in the thesis, the sources can hardly be modeled as Gaussian correlation and the assumption of perfect knowledge of source correlation information is generally invalid. With these limitations, practical DVC designs target at achieving coding efficiency as close to the conventional video coding as possible.

While constructing practical codes, the Wyner-Ziv (WZ) problem can be regarded as a quantizer taking into account of side-information and an SW coder

in tandem as shown in Fig. 2.2.

Figure 2.2: Practical Wyner-Ziv coder.

2.2 Distributed Video Coding

2.2.1 Background

Due to the wide spread of portable wireless communication systems like smart phone cameras and tablet cameras, the demands to relentlessly minimize power consumption and device complexity of video encoder have become one of the major issues in digital data compression and transmission systems. Inspired by the surprising information-theoretic results for distributed source coding (DSC), which provide the possibility of achieving high coding efficiency through shifting the complex source correlation exploitation from the encoder to the decoder, research on low complex video encoding is carried out by conducting the task of exploiting temporal redundancies at the decoder only. By applying DSC together with a transcoding server in the cloud, high power efficiency can be achieved for both mobile transmitting and receiving devices as shown in Fig. 2.3.

Originally targeting at low encoding complexity, the DSC for video source known as distributed video coding (DVC) was initiated in the beginning of this century, and the first work on DVC was reported in 2002 [22] followed by a number of representative DVC schemes [15, 23]. Although great advances have been accomplished in the last decade, the methodologies and techniques on DSC/DVC are not yet mature with some daunting technical challenges. The state-of-the-art

Figure 2.3: Low complex encoding and decoding with transcoder.

DVC algorithms still cannot provide satisfactory or desirable solutions compared with conventional centralized video coding (CVC) schemes. For example, there remains a significant rate-distortion (RD) performance loss of the current DVC systems compared with the CVC standards like H.264/AVC [7]. In this chapter, we will take a divide-and-conquer approach to dissect a typical DVC system and delve into each individual constituent. Through reviewing these elements of DVC, a deeper understanding and more fundamental insights into the RD gap can be obtained.

Apart from the low complexity encoding, more desirable features (e.g., error resilience, coding flexibility and parallel encoding) due to unique natures of DVC mechanism have been explored to greatly enhance video coding and streaming functionality and performance. The related frameworks will be discussed in this chapter.

2.2.2 Advance in Recent Years

As mentioned earlier, the low complex encoding is the original motivation and has been widely accepted as the advantage of DVC. Comparisons about the encoding

and decoding complexity have been provided in [4] as shown in From Table 2.1. The ratios of encoding time of intra and inter H.264 encoding to the most widely used DVC framework in [3] are listed. It can be seen that a much faster encoding can be achieved by DVC due to removing of computational burden from the encoder like motion estimation and rate-distortion optimization.

Table 2.1: Ratio of encoding time between H.264 and DVC system in [3] (reported in [4])

QP	Ratio: Intra / DVC			Ratio: Inter (no motion) / DVC		
	GOP 2	GOP 4	GOP 8	GOP 2	GOP 4	GOP 8
40	1.68	2.82	4.07	1.70	2.91	4.11
39	1.69	2.82	4.12	1.71	2.92	4.17
38	1.68	2.83	4.11	1.71	2.92	4.15
34	1.71	2.87	4.22	1.74	2.95	4.25
34	1.71	2.86	4.22	1.73	2.95	4.24
32	1.71	2.85	4.16	1.74	2.94	4.19
29	1.73	2.90	4.26	1.76	3.00	4.30
25	1.75	2.92	4.32	1.76	2.98	4.29

Such low encoding complexity is achieved by scarifying the rate-distortion performance. In lack of reference information at the decoder, a remarkable rate-distortion performance loss compare with CVC was incurred by DVC initially, but it introduces not only challenges, but also opportunities of this new research paradigm. Through the great research efforts spent on DVC, such gap has been gradually reduced in the past decade. For example, the first generation DVC systems proposed in 2002 [22] demonstrated at least 3dB performance loss against H.263+ inter coding. In 2007, the DVC system named DISCOVER [3] integrated most achievements in the past five years and showed remarkably better performance over H.264 intra coding for low motion sequences like “Mother and Daughter”. However, for higher motion sequences, DISCOVER can only achieve similar performance with H.264 intra coding, and showed more than 1.5 dB loss against H.264 inter coding without motion estimation, where co-located block is directly used as reference. Recently, exciting results have been reported in [24] that when coding with high fidelity (high bit-rate), DVC can achieve comparable

performance (less than 1 dB loss) against H.264 inter coding, while significantly outperforms no motion H.264 inter coding with more than 5 dB for sequences containing large amount of motion regions like “Coast Guard”.

Therefore, it can be expected that the performance gap can be further reduced in the near future. In addition, better performance compared with CVC has already been reported in the presence of transmission channel errors as in [25]. Hence, the research and further improvement on DVC can not only promote low complex encoding, but also other DSC-based emerging applications. The details of DVC and its applications are reviewed in the following.

2.3 Anatomy of Distributed Video Coding

In current DVC system, the video frames are firstly organized into key frames and Wyner-Ziv frames. Key frames are selected every M frames, where M is known as the Group of Picture (GOP) size, and the rest frames are denoted as Wyner-Ziv frames. Each key frame is coded as an intra-frame using conventional video coding standards like H.264/AVC, while the Wyner-Ziv frames are intra-encoded and inter-decoded using DVC approaches. In this section, we describe the key elements and techniques for coding the Wyner-Ziv frames. A block diagram with the key elements for the DVC system is shown in Fig. 2.4.

2.3.1 Practical Distributed Source Coding

Since Slepian and Wolf’s theorem is built upon random binning, the early practical distributed source codes were constructed using the binning method. More specifically, consider independently identically distributed (i.i.d.) binary sequences X and its side-information Y with infinite length. The theorem shows that if all possible codewords of X can be randomly assigned into a sufficient number of bins, the probability is arbitrarily low that more than one codewords in a bin can be found to satisfy the given correlation between X and Y . Note that the number of bins required is determined by the correlation.

Such binning method is illustrated in the following example as shown in Fig. 2.5. Let X and Y be 3-bit binary sequences, and each takes possible values

Figure 2.4: A block diagram for the DVC system.

Figure 2.5: An example of DSC using binning.

from 0,0,0 to 1,1,1. Meanwhile, X is correlated to Y with at most one bit difference. Without loss of generality, we further assume $X = 0,0,1$ and $Y = 0,0,0$. Hence, using the conventional centralized source coding approach, two bits are needed to encode X (indicate the four possibilities of X) given Y at the encoder. Surprisingly, two bits are also adequate to encode X without Y available at the encoder as long as the correlation is known. To achieve so, the eight codewords of X are divided into four bins so that the two codewords in the

same bin are as far from each other as possible in terms of Hamming distance. The index of the bin containing the actual X (Bin with index 01 containing 0, 0, 1 and 1, 1, 0) is transmitted by its index of two bits. At the decoder, X can be identified as the codeword whichever is closer to the side-information Y .

Although binning approach is used for DSC theorem proof and can easily illustrate SW coding process, the simple binning coding described above is low in efficiency in practical SW coding and can cause large decoding error in the case of inaccurate correlation information. More robust and efficient approaches are to recast the DSC problem into a channel coding scenario. In this case, a virtual error prone channel is employed between the binary sequence X and its side-information Y . As a result, the bit differences between X and Y are considered as bit errors caused by the virtual channel. The decoder needs to recover X from the received side-information Y . Such recovery is feasible by applying systematic channel coding to X and transmitting the generated protection bits. Forward error correction (FEC) can be performed by applying the received protection bits to the corrupted sequence Y . Meanwhile, when X and Y are highly correlated with fewer bit differences, which indicates fewer virtual channel errors are introduced, fewer protection bits are required to recover X from its corrupted version Y . Consequently, compression of X is accomplished.

The concept of applying channel coding to DSC was proposed shortly after Slepian-Wolf (SW) theorem. However, practical distributed source codes have been investigated since only around one decade ago, because of development of powerful channel codes and emerging demands of low complex encoder.

The early stage of channel coding based DSC design mainly focused on trellis codes [26]. A number of schemes are developed using the trellis codes [27–31], for Gaussian sources. These channel coding approach can be considered as advanced binning method, where the parity check bits are served as the bin indices. However, the coding efficiency is still relatively low for both symmetric and asymmetric DSC, mainly due to that the correlation between source and its side-information cannot be well exploited.

By realizing the importance of exploiting source correlation, iterative decoding with soft-in-soft-out algorithm has attracted more research interests and become the most popular DSC approach. By using the iterative decoding, the source

correlation is represented as reliability of each received side-information bit, and hence can be better exploited to improve the coding efficiency at the cost of drastically increased decoding complexity. Punctured Turbo codes which transmit only a portion of the generated parity check bits have been developed for both binary DSC [32–36] and DSC with non-binary symbols [37, 38].

Apart from the turbo codes, Xiong’s group proposed low-density-parity-check (LDPC) codes as another powerful tool for DSC [39–41]. Since then, the LDPC codes [42–45], which also employs iterative decoding, have achieved even better coding efficiency than Turbo codes. Although the LDPC codes show superior performance, they still have a disadvantage that different generating matrices must be used for different encoding rates, which leads to a requirement of large amount of storage. To overcome this problem, the rate adaptive LDPC-Accumulated (LDPCA) code has been developed by Girod’s group in [46], which provides compression performance approaching the Slepian-Wolf bound with further advantage that different rates can be achieved without altering the generating matrix of the code. Currently, the LDPCA codes are the most popular Distributed source codes.

Recently, arithmetical codes have been considered as an alternative to the turbo codes and LDPC codes for distributed source coding and attracting increasing research interests [47–53].

2.3.2 Side-Information Estimation

Side-information estimation, also known as decoder motion estimation, has been regarded as one of the most important elements since the first DVC framework, considering the temporal correlation exploitation is the key of video coding. Similar to conventional video coding, the side-information estimation in DVC aims to generate side-information most correlated to current WZ block based on previously reconstructed data. The challenge of the decoder side-information estimation is the lack of current frame information to perform motion search, while the motion search results are required for the decoding of current frame information.

To tackle the problem that the side-information estimation lacks of current frame information to perform motion search, transmitting auxiliary information

was the choice for early DVC systems. In [54], the cyclic redundancy code (CRC) of the quantized symbols is transmitted to verify the DSC decoding result based on each side-information candidate, where the first decoding result matching the CRC is declared as output. Another approach is to transmit robust hash code [55], which is a down-sampled and quantized version of current block, to represent the actual current block for decoder motion search. Unfortunately, these methods require large amount of overhead and can achieve very limited side-information estimation accuracy.

Figure 2.6: Motion compensated extrapolation and interpolation at the decoder side.

Due to these limitations, auxiliary information based decoder motion search

was replaced by decoder motion compensated interpolation/extrapolation [56]. The motion compensated interpolation/extrapolation is based on the assumption that the motion is uniform between the reference frames. Denote current WZ frame as $F(t)$, and two previously reconstructed frames as $F(t-m)$ and $F(t-n)$, where t is the temporal index and m, n with $m < n$ are integers. Note that $F(t-m)$ and $F(t-n)$ can be either intra-frames (key-frames) or reconstructed WZ frames. To perform motion extrapolation for a block in $F(t)$, motion vector $MV = mv_x, mv_y$ is estimated using its co-located block in $F(t-m)$ based on $F(t-n)$. The motion vector is then prorated as $MV_e = MV \frac{m}{n-m}$. Then, the side-information is obtained by applying MV_e to $F(t-m)$. For motion interpolation [57], usually an intra-frame $F(t+k)$ is required. Motion interpolation is usually performed bi-directionally to improve estimation accuracy. To perform motion interpolation for a block in $F(t)$, forward motion vector MV_f is estimated using its co-located block in $F(t+k)$ based on $F(t-m)$. Through prorating, $MV_f^i = MV_f \frac{m}{k+m}$ and the forward side-information is obtained by applying MV_f^i to $F(t-m)$. Similarly, the backward side-information can be obtained by applying $MV_b^i = MV_b \frac{k}{k+m}$ to $F(t+k)$. Examples using adjacent frames are illustrated in Fig. 2.6. A number of schemes have been proposed based on such frame to improve the side-information estimation accuracy, such as filtering based approach [58], mesh based approach [59] and non-linear approaches [60, 61].

Although the motion compensated interpolation/extrapolation has been under development for a few year, the side-information estimation quality is still not satisfactory. The main reason is that the actual motion is not as uniform as assumed. Meanwhile, the motion estimation is carried out without any information of current frame.

Although full information of current frame is not available before completion of all decoding procedures, partial information is usually accessible during decoding. Hence, exploiting such partial information is the key to improve side-information quality. By realizing this, recently, attention has been paid to iterative motion refinement based approaches [24, 62–69]. One category of these approaches investigated motion refinement based on partially decoded erroneous information [62–64, 66, 67]. More specifically, the decoder firstly attempts to decode current WZ frame using received parity check bits and an initial side-information, e.g.,

that generated from motion compensated interpolation/extrapolation. Even if it is very likely that the decoded frame is erroneous due to insufficient parity check bits and low quality side-information, it is used to estimate the side-information for next interaction. It is expected that the errors in the decoded frame can be reduced through each iteration, and thus achieving refined motion estimation results. However, it is difficult to predict the error rate of the decoded frame in each iteration which is related to the amount of transmitted parity check bits. Hence, it is possible to lead to either waste of transmission bit rate or extremely distorted decoded frame which can hardly be used for motion refinement.

To eliminate the rate control problem, motion refinement based on decoded scaled (but correct) current frame [24, 65, 68, 69] has been considered instead of using erroneous frame. Such approaches begin with decoding of a low-resolution version of current frame based on an initial side-information. Then, the motion of higher resolution frame is refined based on motion estimation using the low-resolution information. Following this concept, one of the most remarkable works is the multi-resolution motion refinement (MRMR) scheme proposed by Liu, et.al. [24], which has reported comparable rate-distortion performance as H.264/AVC with high fidelity. More details of the MRMR scheme will be provided in Chapter 4.

2.3.3 Source Correlation Estimation

According to the fundamental DSC theorems, The coding efficiency of the distributed video coding (DVC) systems relies on not only the quality of side-information (SI), but also the accuracy of the statistical correlation information. In other words, the quality of side-information determines the minimum achievable rate bound, while the more accurate source correlation information can help make the DSC performance closer to this bound. Currently, the source correlation is required by the soft-in-soft-out algorithm of the most widely used SW codes including LDPC and Turbo codes described earlier. The source correlation is translated into likelihood estimates for each source bit [32, 39] and then input to these SW decoders. In addition, the correlation estimation is used in the side-information aided quantization/de-quantization including nested scalar

quantizer design [10] and minimum mean square error (MMSE) reconstruction [70]. Furthermore, the correlation estimation can be used for encoder rate control [71, 72] which aims to remove feedback channels of current SW code designs.

In most current DVC systems, the correlation between the current video signal (including frame, pixels and transform coefficients) X and its side-information Y is modeled as zero mean Laplacian distribution:

$$f_Z(z) = \frac{\alpha}{2} e^{-\alpha|z|} \quad (2.2)$$

where $Z = X - Y$ is known as the correlation noise, and $E\{Z\} = 0$ and $\sigma^2 = E\{Z^2\} = 2/\alpha^2$ ($E\{\cdot\}$ indicates the expectation). Since all the above mentioned correlation information like the bit likelihood, MMSE reconstruction offset and encoder estimated transmission rate, can be derived from the Laplacian modeling, correlation estimation actually refers to estimation of the variance σ^2 of the video signal correlation noise Z in current DVC systems [73]. An more accurately estimated σ^2 tends to provide better coding efficiency [74].

A brief history of the correlation noise modeling can be found in [14]. In the early stage of DVC development, the stationary correlation noise is assumed for a whole video sequence. Hence, offline training was suggested to obtain the parameter σ^2 [55, 57]. Considering the time-varying characteristic of video signal, the correlation estimation tends to go from large scale like sequence level to smaller scale levels including frame and coefficient band level, which assumes that stationary correlation noise exists only within a frame or coefficient band. Online stimulation is also applied to further improve the accuracy based on previously decoded frames [73, 75, 76]. Correlation estimation at the coefficient and pixel level has also been studied [13, 74, 77], and better performance has been reported compared with the band level estimation.

However, all these above-mentioned schemes estimate current correlation noise based on previously decoded data, without considering any information from current frame. Like in the side-information estimation step, better estimation accuracy can be expected if partially decoded data of current frame can be exploited. Hence, iterative refinement is also considered for correlation estimation. Recently, a transform-domain adaptive correlation estimation (TRACE) [14] method is

proposed, where the correlation information is updated progressively with more coefficient band decoded.

2.3.4 Quantizer Design

In order to improve the rate-distortion performance of DVC, the design of quantizer should take into account the side-information as suggested by the Wyner-Ziv (WZ) theorem. The quantizer design first considered vector quantization for the case of idea sources [78–82].

In the context of DVC, side-information is generally utilized at the decoder and the correlation information may be required at the encoder for the quantizer design. At the quantization step, uniform scalar quantization is one of the commonly used approaches, using a designed quantization matrices [4] to quantize the coefficients or pixels into 2^M intervals, which can facilitate the bit-plane based Slepian-Wolf coding. To study the non-uniform quantizer design, our work in [83] examined the Lloyd algorithm as well as dead-zone based quantizer design in practical DVC. The most remarkable quantizer design is the study of nested lattice quantization [9] by Xiong’s group, which is based on coset partition, and the code design and the achievable region using nested lattice quantization have been studied in [9–11]. The nested lattice quantization has been proposed for applications such as layered video coding [2, 16] and sensor networks [84]. Our work inspired by the nested lattice quantization is presented in Chapter 3.

On the other hand, the side-information is widely used in the de-quantization step (also known as reconstruction step), which may be different from the approaches in most conventional video compression systems. With assumptions that the side-information available at the decoder is of high quality and the current source and its side-information are highly correlated, the simplest approach is to determine possible values represented by the SW decoded quantization index, and select the value within the possible values that is closest to the side information. Several approaches have been proposed to minimize the mean-squared error based on the estimated correlation between source and its side-information [70, 85, 86].

2.3.5 Transmission Rate Control

Transmission rate control refers to determine the amount of coded SW bits to be transmitted for a successful decoding. The difficulty lies in the lack of side-information at the encoder. To avoid transmitting extra bits, the most widely used strategy in current DVC system is to send coded bits incrementally until successful decoding is achieved. More specifically, at the encoder, the syndrome or parity check bits generated by the Slepian-Wolf codes such as LDPC codes and Turbo codes are buffered and transmitted chunk by chunk. With the received portion of SW codes, the decoder attempts to decode the original bitstream with the help of estimated side-information and correlation information. If the soft-in-soft-out decoding algorithm can converge within a specified number of iterations, successful decoding is declared and the SW decoding process completes. Otherwise, a request of additional coded bits is sent by the decoder through a feedback channel to the encoder buffer.

The limitations of using the feedback channel based rate control strategy is the dramatically increased decoding complexity and latency, since the whole soft-in-soft-out decoding process must be repeated each time additional coded bits are received. As mentioned earlier, the rate control relies on the estimation of correlation noise distribution, based on which the necessary rate to achieve a target decoded quality is calculated. Considering the limited estimation accuracy of correlation noise, hybrid approaches have been developed [87–89], where the encoder starts to transmit certain amount of coded bit chunks together based on an initial estimation, instead of one chunk each time. Hence, the number of feedback can be greatly reduced, and thus the decoding complexity as well as latency. However, in practical scenario, the feedback channel may be totally unavailable, which requires the encoder to be solely responsible for the rate control. Although a lot of efforts have been spent on this area [73, 90–94], the coding efficiency loss is significant due to the poor correlation estimation without side-information at the encoder.

2.4 Emerging Applications of Distributed Video Coding

2.4.1 Error Resilient Video Coding

Since current distributed source coding (DSC) techniques mainly depend on channel codes such as Turbo and LDPC codes, apart from the low complexity encoding, another important application of DSC is to emphasize on error resilient coding. Firstly, schemes were proposed to protect frames with WZ coding concept [95, 96], which transmits a coarsely quantized version of original frames (instead of protecting bits from channel coding) to prevent large transmission errors. Sehgal et al. [1] then attempted to stop error propagation by protecting certain frames periodically against transmission channel errors using WZ coding. Later, our work [97] tackled the error propagation problem using a unified WZ code to protect both virtual channel (between current WZ frames and their reference frames) errors and the real channel errors. Online monitoring the correlation noise and transmission channel noise has been considered in [25] to further improve the error resilience coding efficiency.

Recently, WZ coding has been combined with another error resilient coding paradigm, namely, multiple-description coding [98], in order to tackle the predictive mismatch problem of MDVC. Representative works include the schemes in [99–101]. In [100], a DSC-based MDVC scheme is developed, which attempts to improve the decoding performance using an iterative manner when both the descriptions are received. In [101], reference replacement is considered, where in case one side decoder cannot access its own side information, it can use the corresponding frame of the other description as side information. Compared with the scheme in [101], our work in [102] considers more candidate frames of side-information with the concept of DSC-based flexible decoding described in the next section.

Figure 2.7: DVC for flexible multi-view video decoding.

2.4.2 Flexible Video Decoding

Another interesting development of DVC is the flexible coding for fast random access in both mono-view and multi-view video as presented in [72, 103]. In the conventional predictive video coding system, decoding order is fixed as the encoding order. As a consequence, decoding a certain frame or region requires the decoding of all the reference frames or blocks even they are not requested by the user, which leads to inflexibility and inefficiency in bit-rate and decoding time. On the other hand, in DVC mechanism, the predicted symbol is encoded independently of its reference symbol, and hence any symbol available at decoder can be used as reference instead of a fixed one, thus achieving the flexibility. Flexible video decoding can have a wide spectrum of applications of practical significance. For example, in mono-video coding, Wyner-Ziv frame can be inserted periodically instead of using intra-frame to enable fast access to a frame in a long GOP. In addition, it can be applied in flexible decoding of regions of interest as well. For multi-view video coding, such flexibility is even more desirable for fast switching among pictures corresponding to different views in an efficient way as shown in Fig. 2.7

2.4.3 Parallel Video Encoding

In conventional video coding, serial encoding is usually employed where one frame is encoded after reconstructing its reference frames as shown in Fig. 2.8a. However, with the popularity of mobile camera devices equipping multi-core processors like tablets and smart phones, parallel encoding is desired to achieve faster encoding. Although current CVC architecture allows multi-slice / tile setting for parallel processing of slices / tiles within one frame, coding efficiency loss

Figure 2.8: Serial encoding using CVC vs. parallel encoding using DVC.

is inevitable. On the other hand, in the DVC system, each frame are encoded independently, where the information of reference frames is not required at the encoder. Hence, parallel encoding is naturally enabled by DVC as shown in Fig. 2.8b. Meanwhile, the parallel encoding can be implemented at frame level without introducing additional coding efficiency loss. Some frameworks of DVC based parallel video encoding have been proposed as in [104], and significant encoding speed increase has been reported without hardware acceleration.

Chapter 3

Adaptive Nested Scalar Quantizer for Distributed Video Coding

As stated by the Wyner-Ziv theorem, the quantizer design with consideration of side-information plays an important role in the distributed source coding system. In this chapter, we present our adaptive nested scalar quantizer schemes to improve the distributed video coding performance.

3.1 Introduction

One of the practical quantizer designs with considering side-information in DVC is the coset partition based one-dimensional nested lattice quantizer which use a scalar quantizer followed by a coset channel code [105] (also known as coset partition and binning, and we use the terms “binning” and “coset partition” interchangeably throughout this thesis). A binning scheme partitions the space of all possible code values into disjoint subsets (“bins” or “cosets”), e.g., assigning the code values to the subset in a way to maximize the minimum distance in each subset. In the theoretical proof of the DSC theorems [105], the bins are constructed randomly, and consequently the scheme is characterized in probabilistic terms: the probability that certain received side-information is close to

(or “jointly typical” with) more than one code value in a given bin can be arbitrarily small. This random construction is favorable for theoretical analysis, which, however, is not convenient for practical applications due to its lack of structure. Several structured binning schemes have been developed and widely used in practical DSC and DVC. Starting from Wyner’s first construction of algebraic binning scheme [106], binning schemes using coset codes [107–109] based nested lattice quantization have been investigated. which has been found to be suitable for quantizer design with considering side-information for Wyner-Ziv problem. The indices from the nested lattice quantization can be further compressed using parity check codes such as Low-Density-Parity-Check (LDPC) codes [39, 46] and Turbo codes [32, 35]. The code design and the achievable rate-distortion analysis using the nested lattice quantization have been reported in [9, 10, 17].

The one-dimensional nested lattice quantization (known as nested scalar quantization, or NSQ in short) has been studied theoretically in [10, 17] under assumptions of high bit-rate and ideal source distribution, and applied in [2, 16, 84] for error resilient scalable video coding and sensor networks.

The theoretical analysis of the nested lattice quantization is based on the high bit rate (high resolution) assumption with a precise distribution model of the sources, so that a “good” conventional coset channel code can be constructed to ensure a very small probability that a wrong code value is selected as the decoded result from the received bin, given the side-information. However, in practice, the assumption of high rate may neither hold nor be used as a good approximation. Moreover, there is no accurate statistical model of the correlation between the source and its side-information in most practical scenarios like in distributed video coding. In these practical situations, such a “good” coset channel code is normally not available, leading to a much higher probability that a wrong code value is selected to be the decoded result in a given bin. Thereby, the distortion with MSE metric becomes very large when a wrong code value is selected since the codewords in a bin are far apart from each other. However, we note that the binning method based on the conventional scalar quantizer (SQ) groups the neighboring code values into the same quantization bin, which restricts the range of code values in the bin. In this way, the maximum decoding error given the side-information can be clipped within a certain range even if the correlation

between the source and its side-information is lower than expected. It can be seen that the selection of the conventional nested lattice quantization and SQ-based binning should be made based on the correlation between the source and its side-information to maximize the coding efficiency. Therefore, we propose adaptive 1-D nested lattice quantization schemes to employ the two above mentioned binning methods adaptively according to the correlation estimation, for maintaining their respective advantages while mitigating their limitations.

More specifically, two adaptive schemes are presented. Firstly, we study an adaptive binning scheme by integrating the two coset partition methods based on empirical analysis. We present a practical code design focusing on the adaptive bit allocation in different levels in the transform block coding based on a prediction of correlation noise for minimizing the decoding errors. In the other scheme, we further investigate the nested lattice quantization with more theoretical analysis based on the Laplacian correlation noise modeling in the context of DVC. A correlation threshold is introduced in the selection of the two binning methods to minimize the rate for a given distortion. The threshold can be derived based on a coarse modeling of correlation noise which can be obtained in practical video coding. We apply both the adaptive schemes to distributed video coding and demonstrate superior rate-distortion performance against the non-adaptive schemes.

The remainder of this chapter is organized as follows. In Section 3.2, we provide the fundamentals about the two coset partition methods and discuss their limitations advantages. We present the empirical analysis based adaptive nested quantization algorithm in Section 3.4. Section 3.3.4 describes our further study of the adaptive nested quantization algorithm with Laplacian correlation noise based analysis. The summary of this chapter is presented in Section 3.5.

3.2 Fundamentals of One-Dimensional Nested Quantization

In this section, we will briefly review the theoretical analysis and practical applications of the 1-D nested lattice quantization as shown in Fig. 3.1. As defined in

[10, 17], the nested lattice quantization can be achieved using a scalar quantizer (fine lattice code) followed by a conventional coset channel code (coarse lattice code), as illustrated in Fig. 3.2.

Figure 3.1: An illustration of the 1-D nested lattice quantization.

Figure 3.2: Practical 1-D nested lattice quantization scheme.

Let X be the current source to be encoded by the DSC method, Y be the side-information of X available at the decoder, and $X = Y + Z$, where Z is the correlation noise. To encode the source X using the nested lattice quantizer, the source is firstly quantized into code values (white dots in Fig. 3.2) using a scalar quantizer $Q(\cdot)$ with a stepsize of q . The quantized source is then partitioned uniformly into N bins (N is also known as nesting ratio for the nested lattice) in such a way that the distance between neighboring code values in each bin is $N \cdot q$. In Fig. 3.2, j denotes the index of a code value in a certain bin after the coset channel coding. To achieve compression, the index of the bin which the current

code value belongs to is transmitted instead of the code value itself. Note that the bin indices can be further encoded using entropy coding or another Slepian-Wolf code for more compression. At the decoder, the code value in the received bin which is closest to the side-information Y is selected, and de-quantization is applied to obtain the final decoding result. It can be seen that the decoding process is simpler than the binary Slepian-Wolf codes like Turbo and LDPC codes with iterative belief propagation decoding algorithms.

3.2.1 Theoretical Results on 1-D Nested Lattice Quantization

In [10, 17], a theoretical analysis of nested quantization has been presented. The distortion function based on the MSE metric in the case of high bit rate (high resolution) has been given by

$$D = \frac{q^2}{12} + (N \cdot q)^2 E\left(\left\lfloor \frac{Z}{N \cdot q} + \frac{1}{2} \right\rfloor^2\right) \quad (3.1)$$

where $E(\cdot)$ and $\lfloor \cdot \rfloor$ denote the expectation function and the floor function, respectively. If Z is a zero-mean Gaussian variable with variance σ^2 , (3.1) can be represented as [10, 17]

$$D = \frac{q^2}{12} + 2(N \cdot q)^2 \sum_{j=0}^{\infty} (2j+1) \Phi\left(\frac{N \cdot q}{\sigma} \left(j + \frac{1}{2}\right)\right) \quad (3.2)$$

where $\Phi(\cdot)$ denotes the normal distribution function. From (3.1) and (3.2), it can be seen that the distortion consists of two parts, that is the first term $q^2/12$ known as quantization error from the scalar quantization, and the rest known as binning error due to a wrong selection of the code value in the received bin. The binning error occurs when $|z| \geq N \cdot q/2$. It can be seen that for a given N , a larger quantization stepsize q can reduce the binning error at the expense of increasing quantization error, while a smaller q results in a larger binning error but a smaller quantization error. Therefore, an optimized balance is expected to minimize the sum of the two types of errors with respect to N and q . In [17], an

empirical result on the selection of N and q has been presented to minimize the distortion, that is, $N \cdot q > 5\sigma$ for Gaussian distributed correlation noise.

According to [10], 2-D nested lattice quantizer has better performance than 1-D nested lattice quantizer if the quantization indices are further compressed using an ideal Slepian-Wolf coding. However, [10] also shows that as far as a practical LDPC code based Slepian-Wolf coding is concerned, 1-D nested lattice quantizer performs very close to 2-D quantizer. Therefore, in this chapter we will focus on the discussion for a 1-D nested lattice quantizer for more simplicity in both encoding and decoding process.

The advantages and limitations of the 1-D nested lattice quantizer are illustrated with two special cases where $q = 1$ and $N = 1$, respectively.

3.2.2 $Mod()$ Coset Partition

By setting $q = 1$, the calculation of nested scalar quantization index for the input X is simplified as

$$I_M = X \text{Mod}(N) \quad (3.3)$$

where $Mod(N)$ is the modular arithmetic by N . Hence, we also denote the coset channel coding in the nested scalar quantization as $Mod()$ based binning, and use these two terms interchangeably in this thesis.

With this partition way, if X and Y are highly correlated, the magnitude of noise Z is small and side-information Y is considered “good”. As long as $|Z| < N/2$, there is no error to reconstruct X with the value in the right coset nearest to Y . However, if X and Y are less correlated with $|Z| > N/2$, a decoding error of $N \times \text{round}(Z/N)$ occurs. Note that even in the good case (with high correlation), the decoding error could still be potentially large if insufficient number of bins is used for coding. An example shown in Fig. 3.3 illustrates the two cases, respectively. In each sub-figure, the top line shows a set of M possible values of X with small circles, and the following three lines denote the three cosets ($N = 3$) which partition the set. The actual value small of X is indicated by the solid box, which is the first element in the first coset in the example. The dashed box shows the decoded value with the $Mod()$ coset partition scheme. It can be seen that as the value N increases, the stronger error-free decoding capability can achieve by

the $Mod()$ coset partition scheme at the cost of increasing number of bits.

Figure 3.3: An example of $Mod()$ coset partition with $N = 3$ and $M = 18$.

3.2.3 Scalar Quantizer (SQ) Based Coset Partition

The other extreme case is when N equals to 1. In this case, there is no binning error, leaving only the quantization error. Hence, the coset index for the input X is simplified by

$$I_Q = \lfloor X/q \rfloor \quad (3.4)$$

where $\lfloor \cdot \rfloor$ is the *round* operation. This is equivalent to performing a scalar quantization with the stepsize q . In this way, each coset contains q adjacent values with distance of 1 from each other in order. Consequently, if X and Y are less correlated with $|Z| > q$, this coset partition method has the advantage of clipping maximum decoding error to no more than q . However, even if X and Y are highly correlated with a small $|Z|$ (like $|Z| < q$), there may still be a decoding error of Z . Examples in the two cases are illustrated in Fig. 3.4, respectively.

Figure 3.4: An example of SQ-based coset partition with $q = 6$.

3.2.4 1-D Nested Quantization for Video Coding

The nested quantizer has been applied to video coding and better results have been reported in [2, 16] compared with predictive coding in the presence of transmission errors. In those schemes, the nested quantization is applied to discrete cosine transform (DCT) coefficients. At the encoder, the coefficients are first quantized by the conventional uniform quantizer with a stepsize q , and all the possible quantized code values are partitioned into N cosets of equal size as shown in Fig. 3.2. Only the index of the coset that includes the current quantized coefficient is transmitted. The indices can be converted to bit-planes and encoded by an entropy coder or another Slepian-Wolf coder. At the decoder, with the decoded coset index, the code value can be obtained by selecting the one closest to the side information. The advantage of using the nested quantization is that even if a transmission error E corrupts the side-information, there is still no decoding error as long as $|Z + E| < N \cdot q/2$.

3.3 Distributed Video Coding Using Adaptive Nested Scalar Quantization Based on Empirical Analysis

In this section, we will attempt to apply the integrated coset partition scheme to the practical DVC system, where the bit allocation for the two coset partition methods is achieved adaptively. The adaptive scheme aims to reduce decoding error for a given coding rate. The DVC system using the adaptively integrated coset partition method is illustrated in Fig. 3.6. In our system, video frames are firstly organized as key frames and Wyner-Ziv frames, respectively. The key frames (odd frames) are coded as H.264 intra-frames, while Wyner-Ziv frames (even frames) are coded by the proposed scheme. The adjacent key frames are used to generate side-information for the Wyner-Ziv frame between them through interpolation.

3.3.1 Motivation

It has been shown that when “good” side-information is available (like $|Z| < N/2$), $Mod()$ coset partition is expected to produce error free decoding results. In the case that “good” side-information may not be found, SQ-based coset partition is preferred to limit the maximum decoding error. In view of the two extreme cases, an adaptive coset partition method based on the prediction of correlation noise Z is expected to yield better performance.

Assuming a prediction of the correlation noise, an integration of the two coset partition methods can be developed. Bit allocation between the two coset partition methods is required to strike a good balance between maximizing the error-free decoding capability (by using the $Mod()$ coset partition) and minimizing the maximum decoding error (by the SQ-based coset partition). Given a total of S bits in encoding a coefficient, assume that S_M bits are allocated to the $Mod()$ coset partition method while S_Q bits are for the SQ-based coset partition method ($S = S_M + S_Q$). In this way, the possible decoded values are constrained in an intersection of the two cosets obtained by the two coset partition schemes, respectively. More specifically, it can be obtained that the possible decoded values will

span a range of M/N_Q with the minimum distance of M/N_M between every two values in order, where $N_M = 2^{S_M}$ and $N_Q = 2^{S_Q}$. A general guideline for the bit allocation is to maximize S_M for error-free decoding while limiting the possible binning errors by increasing S_Q , depending on whether good side-information is available (small correlation noise) or not (large correlation noise). In the following, we will discuss a practical way for the prediction of the correlation noise in the context of DVC.

Consider two adjacent frames in DVC, where one is coded using DSC approach (known as Wyner-Ziv frame), and the other one is used as the side-information (known as key frame). A lower motion region normally corresponds to a smaller correlation noise Z , for which “good” side-information can be obtained from the reference frame at the decoder side. On the other hand, a higher motion or occlusion region is associated with larger Z , which means “good” side-information may not be found. Since motion estimation is not applied at the encoder, the value of sum of absolute differences D (denoted as co-difference for simplicity) between a block to be encoded and its co-located block in the reference frame is used to predict and model the correlation noise at the decoder side, which is the residual between the side-information obtained through decoder motion compensation and the original Wyner-Ziv frame. Fig. 3.5 shows the relationship between co-difference and the correlation noise at the decoder side in both spacial and transform domain, where the correlation noise is obtained as the sum of absolute differences between a 8×8 block to be coded and its best-matched block based on the full search in a 15×15 window. From the figure, we can see that the co-difference matches well in general the correlation noise as a whole despite of some large variances in the case of large co-differences in both spacial and transform domain. It can also be seen that when the co-difference tends to be large, the correlation noise becomes more difficult to predict, and our adaptive scheme is designed to clip the error caused by inaccurately predicted correlation noise. The scheme presented in this section the block-wise co-difference in DCT domain is employed to predict the correlation noise.

(a) Spacial domain

(b) Transform domain

Figure 3.5: The relationship between averaged co-difference and the correlation noise at the block level for four different sequences.

Figure 3.6: Proposed distributed video coding scheme.

3.3.2 Encoder

3.3.2.1 Discrete Cosine Transform and Quantization

It is known that DCT is widely adopted for compression due to its good energy compaction property. In our DVC scheme DCT is applied as well to each $n \times n$ block in a Wyner-Ziv frame. In our experiments, $n = 8$ and different dead zone quantizers are used for varying rate-distortion performances.

3.3.2.2 Mode Classification For Bit Allocation

Different modes are designed for different predicted block correlation noise, which are based on the co-difference values in DCT domain. The quantized DCT coefficients for a Wyner-Ziv block and its co-located block are denoted as $C(i, j)$ and $C_R(i, j)$, respectively. Then the block distance in the DCT domain, known as weighted co-difference (WD), is calculated as:

$$WD = \sum_{i=0}^{n-1} \sum_{j=0}^{n-1} (|C(i, j) - C_R(i, j)| \times W(i, j)) \quad (3.5)$$

where $W(i, j)$ is the weight factor for the (i, j) component, where $i, j \in \{0, 1, \dots, n-1\}$, and $(0, 0)$ indicates the DC component. The weight generally decreases from DC coefficient to higher frequency coefficients. According to our simulations, the overall coding efficiency is not sensitive to a small variation of the weights

selection. In our simulation, we select the weight as:

$$W(i, j) = \begin{cases} 1/1.5^{i+j} & \text{if } i + j \leq 4 \\ 0 & \text{elsewise} \end{cases} \quad (3.6)$$

Based on the value of WD , we can classify the Wyner-Ziv block into several different modes for bit allocation. Three levels of bit allocation are considered as follows:

(i) Level 1 - bit allocation (block level) per Wyner-Ziv block. Different Wyner-Ziv blocks may be assigned with different amount of bits according to the predicted block correlation noise.

(ii) Level 2 - bit distribution to each coset partition method for a Wyner-Ziv block, given the bit budget for the block determined in Level 1.

(iii) Level 3 - bit allocation (coefficient level) for each quantized DCT coefficient in a Wyner-Ziv block, after Level 2 assigns a certain number of bits to the block.

Specifically, we design the three level bit allocation as follows. Empirically, eight modes are used to distinguish the predicted correlation noise of each 8×8 Wyner-Ziv block, and the associated three levels of bit allocation are given in Table 3.1-3.3. In Table 3.1, if a block is classified to one of Mode 1,2,3, the co-located block is directly used as side-information in view of the smaller WD values, thus saving the rate for coset indices and computation for motion search in the decoder. Zero bit for Mode 1 means that no bits will be consumed for Wyner-Ziv blocks classified to Mode 1 except the mode information as overhead. In this case, the decoder will directly reconstruct the Wyner-Ziv blocks by copying the co-located block in the key frame. Note that based on the bit allocation in percentage shown in Table 3.3, if the calculated number of bits for some coefficient is 1 or 0, we will not assign any bits for encoding the coefficient and the decoder will reconstruct the coefficient value by copying the corresponding one from the side-information block. This is because there is little improvement if the possible values are partitioned into only two cosets with 1 bit assigned.

Table 3.1: Level 1 - bit allocation per Wyner-Ziv block.

Mode	WD	Number of bits
1	< 1	0
2	< 8	20
3	< 18	25
4	< 24	30
5	< 32	42
6	< 40	66
7	< 50	118
8	≥ 50	133

Table 3.2: Level 2 - bit allocation in percentage to the two coset partition methods.

Mode	$Mod()$ partition	SQ Partition
1	100%	0%
2	100%	0%
3	100%	0%
4	66.7%	33.3%
5	33.3%	66.7%
6	0%	100%
7	0%	100%
8	0%	100%

3.3.2.3 Encoding With The Adaptive Coset Partition Method

Each quantized coefficient $C(i, j)$ is encoded using the proposed adaptively integrated coset partition method based on the above bit allocation design. The obtained indices from different coset partition methods for different coefficients will be further coded using Slepian-Wolf code (LDPC code) as in [2, 17] for further exploiting the correlation among the indices. Note that the mode information for each Wyner-Ziv block from the mode classification is also included as overhead, which can be further entropy coded using arithmetic codes. The percentage of the overhead information in bit-rate typically decreases as overall bit-rate increases, and can be less than 5% for high bit-rates.

Table 3.3: Level 3 - bit allocation in percentage for each DCT coefficient $C(i, j)$ per block.

$i + j$	Percentage allocated
0	20
1	7.5
2	5
3	5
4	3
5	2.5
≥ 6	0

3.3.3 Decoder

3.3.3.1 Identifying Possible Values

In the decoder, Slepian-Wolf decoding is performed to reconstruct the coset indices, and we need to reconstruct each DCT coefficient by decoding its coset indices. Consider a coefficient $C(i, j)$ in a Wyner-Ziv block, for which two coset indices, $I_M(i, j)$ and $I_Q(i, j)$, are obtained by the two partition methods, respectively. For the $Mod()$ partition scheme, the coset $\{C_M(i, j)\}$ of index $I_M(i, j)$ contains $M(i, j)/N_M(i, j) = M(i, j)/2^{S_M(i, j)}$ values. On the other hand, for the SQ-based partition, the coset $\{C_Q(i, j)\}$ with $I_Q(i, j)$ includes $M(i, j)/N_Q(i, j) = M(i, j)/2^{S_Q(i, j)} = T(i, j)$ values, where $T(i, j)$ is the stepsize, and $M(i, j)$ is the number of possible values for $C(i, j)$, and $S_M(i, j)$, $S_Q(i, j)$ are the number of bits used for the two coset partition methods, respectively. Based on (3.3) and (3.4), the values contained in the two cosets can be obtained respectively as:

$$\{C_M(i, j)\} = I_M(i, j) + k \times N_M(i, j) \quad (3.7)$$

$$\{C_Q(i, j)\} = I_Q(i, j) \times T(i, j) + m \quad (3.8)$$

where $k = 0, 1, \dots, M(i, j)/N_M(i, j) - 1$ and $m = -\frac{T(i, j)}{2}, \dots, 0, 1, \dots, \frac{T(i, j)}{2} - 1$

By taking the intersection of the two cosets, a subset $\{C_{PQ}(i, j)\} = \{C_M(i, j)\} \cap \{C_Q(i, j)\}$ is identified, which contains $M(i, j)/(N_M(i, j) \times N_Q(i, j))$ possible values and one

of them will be selected as the decoded value for $C(i, j)$ in the following steps.

3.3.3.2 Finding Side-Information Block

The motion search is carried out to identify the “best” side-information block. In this chapter, we consider the motion compensated interpolation (MCI) [110, 111] to generate the side information for the DVC system, which is shown to strike a good balance among the transmission rate, side-information quality and decoding complexity.

3.3.3.3 Reconstructing Wyner-Ziv Block

Once the side-information block is found, the same DCT and quantization is applied to the block. From the possible value subset $\{C_{PQ}(i, j)\}$, we can determine the reconstructed value for each encoded DCT coefficient, which is selected to be the closest to the corresponding coefficient in the side-information block. Then inverse-DCT is performed to obtain the final reconstruction of each block.

3.3.4 Experimental Results

Experiments were carried out to compare the proposed DVC scheme against the relevant methods. Two CIF video sequences (“Foreman” and “Mother and daughter” at 15fps) containing low to high motion were tested. In the experiments, the odd frames were intra-coded as key frames, while the even frames were coded using different Wyner-Ziv approaches. Only the coding results of the Wyner-Ziv frames are shown in the following comparisons.

The rate-distortion comparison of different coset partition methods on the two video sequences is shown in Fig. 3.7, where the results by the state-of-the-art centralized video coding standard H.264 are also included. Arithmetic coding was used to code mode information of the Wyner-Ziv blocks. Note that the scheme with $Mod()$ coset partition method followed by the LDPC code is equivalent to the Slepian-Wolf coded NSQ (SWC-NSQ) scheme in [2, 10, 17]. From this figure, we can see that the proposed scheme outperforms the other two coset partition based methods although there is still a remarkable performance gap from the H.264.

(a) "Foreman" (15fps)

(b) "Mother and Daughter" (15fps)

Figure 3.7: Rate-distortion comparisons (H.264 baseline codec JM12.4 is used with RDO off).

(a) "Foreman"

(b) "Mother and Daughter"

Figure 3.8: Frame based PSNR comparison of different coset partition methods.

Moreover, Fig. 3.8 compares the frame based PSNR values by the three coset partition based DVC schemes for the video sequence "Foreman" and "Mother and

Figure 3.9: Visual quality comparison of different coset partition methods (foreman.cif, frame 180).

daughter”, respectively. For a fair comparison, all the coset partition methods were applied with the same bit-rate at both block and frame levels. To compare the actual performance of the three coset partition methods, the optimized reconstruction in [70] is not employed to the reconstructed Wyner-Ziv frames. It can be seen that the *Mod()* coset partition scheme suffers with the lowest PSNR values in Wyner-Ziv frames containing occlusion and relatively larger motion areas, e.g., from Frames 175 to 185 (camera panning) of “Foreman” and around Frame 12 (Mother’s hand appearing and moving fast) of “Mother and Daughter”. However, it performs well in other frames with no or low motion as expected. The performance of the SQ-based coset partition scheme degrades due to low motion regions in almost all frames. In contrast, our scheme performs as well as the *Mod()* method for the frames with no/low motion, while avoiding large distortion

in the frames with occlusion or large motion.

A comparison on visual quality of the reconstructed images by different coset partition methods is shown in Fig. 3.9. The selected frame (Frame 180 of foreman) contains relatively high motion due to camera panning, which fails the $\text{Mod}()$ method in SWC-NSQ with large decoding error (e.g., the bright blocks in Fig. 3.9b). On the other hand, the SQ-based method can restrict a large decoding error but may suffer constantly small errors, with lower PSNR compared with the proposed scheme. The proposed scheme shows the best visual quality of the reconstructed image in this instance.

3.4 Distributed Video Coding Using Adaptive Nested Lattice Quantization and Code Design Based on Laplacian Correlation Noise Modeling

Although we have demonstrated the improved coding efficiency by using the adaptive scheme in the previous section, more analytical results and discussions are presented in this section. Compared with scheme described in the previous section, in this section we focus on using two modes (namely, the conventional nested 1-D (*Mod* binning) lattice quantization and the SQ-based binning) adaptively to saving the overhead bit-rate used for mode index. Meanwhile, further analysis is provided by applying the theoretical results from [9–11] to practical video coding. In addition, instead of using the previously described empirical correlation noise prediction method, we consider the widely used Laplacian correlation noise modeling.

3.4.1 Limitations of Nested Lattice Quantization Based on Laplacian Correlation Noise Modeling for Video

As mentioned in [2], due to the coset channel code employed in the coding process, the nested 1-D lattice (scalar) quantizer scheme suffers a large binning error in

the case of $|Z| \geq N \cdot q/2$, where the number of bins N is related to the bit rate and q determines the quantization distortion. For a given q , which determines the quantization error, reducing the bit rate related to N is conflicting with enlarging the distance $N \cdot q$ between neighboring code values in a bin to reduce the binning error. Therefore, an appropriate N needs to be found to strike the best trade-off.

Although some studies have been carried out in [17] and some empirical results have been presented to address the optimization problem for the Gaussian correlation noise at high resolution, there is still no any good solution to the practical distributed video coding where the correlation noise Z may not be accurately modeled, especially in the region of large magnitude of Z . From the results in [13, 112], the temporal correlation noise for video sequences in pixel domain can be approximated as a Laplacian distribution

$$p(z) = \frac{\alpha}{2} e^{-\alpha|z|} \quad (3.9)$$

The parameter α is calculated based on the variance at sequence or frame level, and the belief propagation decoding algorithm used in Turbo or LDPC decoding will utilize the correlation model to achieve good performance. Some online and offline training techniques have been given in [13] to make the modeling more accurate in practice. An example can be found in Fig. 3.10a for a relatively good Laplacian approximation with $\alpha = 0.35$ (sequence level) can be achieved for modeling the pixel domain temporal correlation noise of “foreman.qcif” sequence. However, this approximation is still not accurate enough to justify using (3.1) for determining optimal q and N , especially in the case of low to medium coding rate. Since a large $|Z|$ with $|Z| \geq N \cdot q/2$ can result in remarkably large binning error compared with the quantization error, decoding error will increase drastically if the approximation model is used to determine optimal N and q which may ignore some large correlation noises. As shown in Fig. 3.10b with a closer observation, the probability of a large correlation noise value such as $|Z| \geq 20$ is much higher than that of Laplacian approximation. Table 3.4 also shows the substantial gap between the actual probability and the approximated one for the “foreman.qcif” sequence. For example, according to Table. 3.4, if the Laplacian approximation is used, the correlation noises with magnitudes larger than 60 can be ignored since

the probability is too low and nearly binning-error-free decoding can still be achieved with $N \cdot q/2 = 60$. However, the probability with these large correlation noises is much higher and much larger binning errors will be introduced if they are ignored when selecting N and q (it may lead to bright or dark pixels to degrade the perception quality significantly according to our experiments). There will be no compression from Slepian-Wolf coding if we choose $N \cdot q = 256$ to eliminate the binning error of 8-bit image. The 1-D nested lattice quantization faces the same problem of inaccurate modeling for block-wise correlation noise and in frequency domain, which may cause more destructive effect to the block reconstruction. Therefore, the lack of an accurate correlation noise model with the consequently large binning error in sequence impede the wide applications of the 1-D nested lattice quantization, especially in distributed image and video coding, where the content can not be modeled well.

Table 3.4: Statistics comparisons of the actual correlation noise Z against the Laplacian approximation with $\alpha = 0.35$ for “foreman.qcif” sequence.

$ Z $	Actual probability	Laplacian probability
≥ 20	9.17×10^{-2}	3.19×10^{-4}
≥ 40	3.18×10^{-2}	2.91×10^{-7}
≥ 60	1.3×10^{-2}	2.65×10^{-10}
≥ 100	1.4×10^{-3}	2.21×10^{-16}
≥ 128	1.51×10^{-4}	1.22×10^{-20}

3.4.2 Proposed Adaptive Nested Quantization

In view of the major obstacle that impedes practical applications of the 1-D nested lattice quantization, we propose a simple yet effective scheme to select between the conventional nested 1-D lattice quantization and the scalar quantization (SQ) based binning adaptively according to the correlation noise Z , as illustrated in Fig. 3.11. A threshold T is introduced in the selection between the two methods. Specifically, when $|Z| < T$, the conventional nested 1-D lattice quantization is applied. Otherwise, the SQ-based binning is applied.

Considering the respective advantages and limitations of the two binning methods discussed above, we propose applying them adaptively based on the value

(a) a global view

(b) a closer observation

Figure 3.10: Approximation of correlation noise Z .

Figure 3.11: Proposed adaptive 1-D nested quantization.

of correlation noise. In the following section, we will determine a good threshold T for the selection between the two binning methods, based on a coarse correlation noise model. Note that the SQ-based scheme is similar to the approach used in distributed video schemes [2, 15], which discards or intra-codes the low correlated bit-planes and performs refinement with side-information. However, none of these existing schemes has made quantitative studies on the threshold which determines whether a bit-plane should be distributed source coded or discarded/intra-coded, while only very coarse guiding rule is given. Our work in this chapter attempts to make such a quantitative investigation for the threshold determination to maximizing the coding efficiency.

Our empirical study shows that in the practical video coding, the distribution of $|Z|$ cannot be accurately modeled, especially in the regions with large $|Z|$ values, which makes it difficult (if not impossible) to find an appropriate q for small or no binning error. Even a small portion of pixels with $|Z| \geq N \cdot q/2$ can result in a large reduction in PSNR thus degrading the reconstruction quality significantly. One way to eliminate this problem is to ensure $N \cdot q/2$ being larger than the maximum value of $|Z|$ with either a large N or a large q , which suffers from very little compression gain or a large distortion, respectively. A better way may be using multiple modes with different $N \cdot q$ values for different correlation regions. However, without an accurate distribution model for the correlation noise, the thresholds for the selection among the multiple modes are very difficult to be determined. Moreover, the overhead information will increase if more modes are

involved. In view of these considerations, our proposed scheme is more practical to address the binning error problem by selecting only the coset channel coding ($Mode()$) and the SQ-based binning adaptively.

3.4.3 Code Design

Figure 3.12: Distribution of block-wise correlation noise $|Z|$ for “foreman.qcif” sequence.

Using the decoder motion compensated interpolation/extrapolation scheme, we consider statistical characteristics of the block-wise correlation noise $|Z|$, which is the largest difference between co-located pixels in the current block and its side-information block. As an example, probability density function of the correlation noise $|Z|$ for “foreman.qcif” sequence with $GOP = 2, 4, 8$ are shown in Fig. 3.12. From the figure, it can be seen that similar to the situation at pixel level, the block-wise Laplacian approximation is not accurate enough to apply the theoretical analysis results in [17]. However, the approximation may be applied in our adaptive scheme without producing large binning error. Although the two

regions cannot be separated optimally because of the lack of accurate correlation noise modeling, the decoding error due to unexpectedly large correlation noise can be suppressed using the SQ-based binning which makes our scheme practical. In the following we will determine the threshold T .

We aim to find the threshold T that minimizes the rate for a given target distortion which is determined by the quantization stepsize q used for the key frames as well as the Wyner-Ziv frames in this chapter (to achieve constant reconstruction quality for the source and its side-information). To apply the coset channel coding scheme in the case of the correlation noise $|Z| < T$, it requires that $T \leq N \cdot q/2$ for free binning error. Therefore, for a given q , we have $N \geq 2T/q$, resulting in a minimum index rate of $R_{Coset} = \log_2(2T/q)$ without considering entropy coding. On the other hand, if $|Z| \geq T$, the SQ-based method is selected, and the index rate required is $R_{SQ} = \log_2(M/q)$ without entropy coding, where M is the number of possible values of the source and M/q is the number of bins. In the context of 8-bit pixel domain distributed video coding, we have $M = 256$ possible values. In view that there are static background blocks in video sequences, regions with zero quantized correlation noise, i.e., $|Z| < q/2$, can be directly reconstructed from the motion compensated side information to simplify the coding process. Therefore, we use one special mode of “skipping” to represent those blocks, which does not require transmission of the coset indices, i.e., $R_{Skip} = 0$), thus saving the index rate. Note that the three coding methods (skipping mode and the two binning methods) without binning errors produce the same distortion for a block, which depends on the quality of the side-information (or equivalently its corresponding quantization stepsize q). In this sense, the rate minimization for a given q is de facto rate-distortion optimization. Consequently, for a given q , the average index rate per pixel with the adaptive binning/coding schemes can be calculated as

$$R = Pr(Skip) \times R_{Skip} + Pr(Coset) \times R_{Coset} + Pr(SQ) \times R_{SQ} \quad (3.10)$$

where $Pr(Skip)$, $Pr(Coset)$ and $Pr(SQ)$ indicate the probabilities by selecting skipping mode, the coset channel coding and the SQ-based binning for a block,

respectively. Apart from the index rate shown in (3.10), overhead of indicating the choice of a coding method for each block is also needed. In view that the overhead on total bit rate is small in terms of percentage with respect to the index rate, e.g., about 4% at the index rate of 1 bpp or even smaller at higher rates according to our experiments, the averaged index rate in (3.10) can be regarded as a good approximation of the total average rate per pixel. With the Laplacian approximation as shown in (3.9), the probability of skipping mode corresponding to zero quantized correlation noise is

$$Pr(Skip) = Pr(|Z| < q/2) = 1 - e^{-\alpha q/2} \quad (3.11)$$

Then the probability of using the coset channel coding becomes

$$\begin{aligned} Pr(Coset) &= Pr(q/2 \leq |Z| < T) = 1 - (1 - e^{-\alpha q/2}) - e^{-\alpha T} \\ &= e^{-\alpha q/2} - e^{-\alpha T} \end{aligned} \quad (3.12)$$

and the probability of using the SQ-based binning scheme is

$$Pr(SQ) = Pr(|Z| \geq T) = e^{-\alpha T} \quad (3.13)$$

Hence, (3.10) becomes

$$\begin{aligned} R &= (1 - e^{-\alpha q/2}) \times 0 + (e^{-\alpha q/2} - e^{-\alpha T}) \log_2(2T) \\ &\quad + 8e^{-\alpha T} - \log_2 q \\ &= e^{-\alpha q/2} \times \log_2(2T) - e^{-\alpha T} \log_2(2T) + 8e^{-\alpha T} - \log_2 q \end{aligned} \quad (3.14)$$

To minimize the rate, we can differentiate R with respect to T yielding

$$\begin{aligned} \partial R / \partial T &= e^{-\alpha q/2} \frac{1}{T} \log_2 e - e^{-\alpha T} \frac{1}{T} \log_2 e \\ &\quad + \alpha e^{-\alpha T} \log_2 2T - 8\alpha e^{-\alpha T} \end{aligned} \quad (3.15)$$

It is difficult to obtain a closed-loop expression of T by solving $\partial R / \partial T = 0$. Alternatively, T can be estimated numerical way. Fig. 3.13 plots the rate R in (3.14) as a function of T with varying q 's and α 's. From the plot, we can see that R reaches the minimum at different T 's with varying α 's and q 's. With more

simulations using different α 's, we obtain the optimal T 's as a function of α for different q 's shown in Fig. 3.14a. In Fig. 3.14b, We also plot $N = 2T/q$, which is the minimum number required for error-free binning in the coset channel coding. From Fig. 3.14a and Fig. 3.14b, it can be seen that, for a given q , the optimal T and N are not very sensitive to a small variation of α when $\alpha \geq 0.3$ since the curves tend to be flat. Fig. 3.14 can serve as a guideline in the selection of T and N with different α and q values.

With the threshold obtained, we can apply this scheme to distributed video coding. In view that the main purpose is to compare the proposed adaptive nested quantization scheme against the conventional DVC scheme without nested quantization in terms of coding efficiency, we consider the two distributed video coding schemes in pixel domain for simplicity. The correlation noise is assumed to be known at the encoders for both schemes so that the minimum rates can be expected respectively. The details of the DVC system with our adaptive scheme are as follows.

3.4.4 Implementation of the Adaptive Nested Quantization Scheme

Similar to our first scheme, video frames are firstly organized as key frames and Wyner-Ziv frames, respectively. The Wyner-Ziv frames are coded by the proposed scheme. We use the LDPCA codes proposed in [46] to further compress the indices generated by the nested quantization. Different from the first scheme, the two methods are selected based on the threshold derived above.

We assume the encoder knows the correlation noise Z in encoding which is possible in practical scenarios. For example, in the case of a feedback channel being available, this can be done by signaling the decoder motion search result to the encoder via the feedback channel. Even without a feedback channel this assumption may still be valid in some practical applications such as DSC-based error resilience coding in [1, 2, 97] where motion search can be done at the encoder (note that the DSC there is for error resilience purpose instead of light-weight encoding). Moreover, in other distributed video coding schemes, the correlation noise may also be estimated or predicted. We would highlight that our work in

(a) R versus T for four different α 's (0.04, 0.08, 0.16, 0.32) and two q 's (16 and 32)

(b) R versus T for five different q 's (4, 8, 16, 32 and 64) and two α 's (0.08 and 0.3)

Figure 3.13: Rate R in (3.14) as a function of T with varying q 's and α 's.

(a)

(b)

Figure 3.14: Plot of optimal T and N as a function of α with varying q 's.

this chapter is to compare the proposed scheme against the conventional DVC scheme without nested quantization in terms of coding efficiency, assuming that

the correlation noise for each block is known at the encoder. With this assumption, the two schemes are fairly evaluated by themselves without considering other “external” factors. We would point out that the knowledge of correlation noise for each block at the encoder is only used for the selection of a binning method in our scheme. Note that the coarse correlation noise distribution mode for a video sequence instead of the exact correlation noise for each block is used for rate-distortion optimization, the latter of which may introduces high computational complexity and large amount of overhead, In fact, the assumption of knowing the accurate correlation noise at the encoder may be loosen with an estimated correlation noise, in view that an inaccurate correlation noise distribution model is employed in our scheme.

In our scheme, the threshold T can be determined with the help of Fig. 3.14a, which is the numerical solution to minimizing the rate in (3.14). Each non-zero block-wise (quantized) correlation noise is compared with the threshold T , based on which one of the two binning schemes will be selected accordingly. If the coset channel coding is selected, the number of bins N is obtained as $N = 2T/q$ as shown in Fig. 3.14b. The indices generated by the coset channel coding are Slepian-Wolf coded again using LDPC codes as suggested in [17] to further reduce the rate. If SQ-based binning method is selected, $256/q$ bins are used for the 8-bit pixel domain DVC. The indices generated by the SQ-based binning are further coded using arithmetic coding due to the low correlation to the side-information. The overhead information of selecting a coding method for each block is also coded using arithmetic coding.

On the other hand, for the existing DVC scheme without nested quantization, the crossover probability between bit planes in a Wyner-Ziv frame and a key frame can be obtained from the correlation noise known at the encoder, based on the same assumption discussed above for a fair comparison. In this way, the coding rate of the LDPC code can be accurately determined and the best coding efficiency can be achieved for the conventional DVC scheme. That is, the minimum rate is used to achieve error-free LDPC coding with the quantization distortion only. Note that the skipping blocks with zero quantized correlation noise are directly reconstructed from the side-information for both the testing schemes.

3.4.5 Experimental Results

In this section, we compare our proposed scheme against the conventional DVC scheme without nested quantization in terms of coding efficiency.

We select three typical video sequences with mixed motion (“Foreman.qcif” of 400 frames), low motion (“Mother & daughter.qcif” of the first 400 frames) and high motion (“Stefan.qcif” of 296 frames) in the experiment. In order to determine the N and T to be used in our scheme, we also plot the block-wise distributions of correlation noise with different GOP sizes (2, 4, 8) for “Mother & daughter.qcif” and “Stefan.qcif” sequences respectively in Fig. 3.15. The α values of the Laplacian approximation to the distribution are estimated experimentally, e.g., 0.08 for “Stefan” and 0.3 for “Mother & daughter”. It can be seen that the parameter α tends to be larger in low motion sequences like “Mother & daughter.qcif” sequence, becoming smaller for high motion sequences like “Stefan.qcif”, where a large correlation noise appears more likely. It can also be seen that the correlation noise distributions for different GOP sizes are close to each other, which implies the GOP size may not affect the selection of T and N in our scheme. It is also observed from Fig. 3.15b that a good Laplacian approximation to the distribution for some video sequences like “Stefan.qcif” may not exist. Fortunately, performance of the proposed scheme will not degrade in that case, which will be shown in the following.

Now we discuss the selection of N based on the numerical solution to minimizing the rate R in (3.14) as shown in Fig. 3.14b. Table 3.5 lists the selected N providing the best results with respect to different q values for the three testing sequences. In our experiments, we find that as long as N is selected to be an odd number around the value suggested in Fig. 3.14b, the similar rate-distortion performance can be obtained. The reason for choosing an odd number of N is to avoid an ambiguous decoding when two code values have the same distance to the current side-information.

Fig. 3.16 illustrates the rate-distortion comparison results by the two testing DVC schemes in coding the Wyner-Ziv frames with different GOP sizes (2, 4, 8) for “foreman.qcif”, “Mother & daughter.qcif”, and “Stefan.qcif” sequence, respectively. From the experimental results, it can be seen that based on the Laplacian

(a) "Mother & daughter.qcif"

(b) "Stefan.qcif"

Figure 3.15: Correlation noise distribution with different GOP sizes of 2, 4 and 8.

Table 3.5: Selection of N with respect to q for different sequences

Sequence	q	Selected N	N based on Fig. 3.14b
“Foreman” $\alpha = 0.08$	4	7	9
	8	5	6
	16	3	4
	32	3	3
	64	3	3
“Mother & daughter” $\alpha = 0.3$	4	5	5
	8	3	4
	16	3	3
	32	3	3
	64	3	2
“Stefan” $\alpha = 0.08$	4	7	9
	8	5	6
	16	3	4
	32	3	3
	64	3	3

distribution approximations, which coarsely model the correlation noises, our proposed scheme can achieve significant coding gains over the conventional DVC scheme without nested quantization, especially in the more meaningful region with PSNR larger than 25dB. More specifically, with the same distortion, our scheme can reduce the bit rate by up to nearly 20% for “Stefan.qcif” in the case of a large GOP size (4 or 8).

In the low bit-rate region, the proposed scheme performs closely to the conventional DVC scheme. The reason is with a large quantization stepsize, the remaining bit-planes are those most significant ones that are very highly correlated to their side-information, leading to very good LDPC coding performance. As mentioned before, there may not be a good Laplacian approximation to a distribution in some situations. However, even with a relatively poor approximation, which may lead to some deviations in selecting N and T , a similar coding gain can still be obtained using the proposed scheme as shown in Fig. 3.16, where up to about 18%, 15% and 20% bit rate reductions are achieved for “Foreman”, “Mother & daughter” and “Stefan”, respectively. This shows our scheme is robust to the accuracy of the correlation noise distribution model, with only a minor cost

of rate as long as no big and frequent deviations occur in the selection of binning methods.

Note that our work in this chapter is to fairly compare the proposed scheme against the conventional distributed video scheme in a same domain, where relative performance comparison (rather than the absolute compression ratio) of the two schemes is our concern. For implementation simplicity, pixel domain is therefore considered, which normally leads to lower compression ratio for both the testing schemes. The same consideration was taken in the pixel domain based distributed video coding scheme in [113], which was reported to produce similar high bit rates as ours.

3.5 Summary

As one of the distributed source coding quantization techniques, one-dimensional (1-D) nested lattice quantization using a scalar quantizer followed by a coset channel code ($Mod()$ based binning) has been theoretically studied with the assumptions of high resolution and Gaussian correlation noise. However, in many practical applications, the high resolution assumption may neither holds nor be used as a good approximation due to a constraint on bit rate, while there may not exist an accurate statistics model for the correlation between the source and its side-information, both of which render the theoretical results inaccurate or even invalid. In the practical situations like video coding where the correlation noise cannot be modeled or predicted precisely, the channel coset code in the nested quantization is not always desired due to a large reconstruction error introduced when the correlation noise goes beyond the error correction capability of the coset channel code. In contrast, the scalar quantization (SQ) based binning has an advantage that the maximum decoding error can be clipped in a certain range even though there is an unexpected large correlation noise.

In this chapter, two adaptive 1-D lattice quantization schemes for distributed source coding are presented to take advantage of their respective strengths while circumventing their weaknesses. The first scheme integrates the two coset partition methods to minimize the decoding errors, where a mode design for bit allocation between the two schemes is developed based on prediction of the cor-

relation noise. On the other hand, the second scheme attempts to provide more analysis and minimize the rate for a given distortion, which employs the two binning methods adaptively according to a correlation threshold. An effective approach to determining the threshold is developed based on a coarse correlation noise model. We apply the proposed scheme to video coding and compare it against the conventional non-adaptive DVC schemes. The experimental results show that the proposed schemes outperform the competing schemes significantly in terms of rate-distortion performance.

(a) "Foreman.qcif"

(b) "Mother & daughter.qcif"

(c) "Stefan.qcif"

Figure 3.16: Rate-distortion performance comparison of the proposed scheme against conventional DVC scheme with different GOP sizes of 2, 4 and 8.

Chapter 4

Exploiting High-Order Correlation in Distributed Video Coding

To further improve the distributed video coding performance, research on efficient correlation exploitation is one of the key directions. In this chapter, we present our scheme to exploit the high-order correlation in distributed video coding to achieve better performance than existing first-order correlation exploitation schemes.

4.1 Introduction

As described in previous chapter, there is still a remarkable performance gap between CVC and DVC in terms of rate-distortion. A substantial research effort in DVC is therefore focused on improving the compression efficiency. One of the most important approaches to increase coding efficiency is to improve the SW coding performance through correlation noise modeling between a source and its side-information [12–14]. In the context of DVC, the correlation noise is defined as the difference between the current frame and its side-information frame estimated through decoder motion estimation. Such correlation noise is typically modeled as a Laplacian distribution, which is used to extract the bit a priori probability estimation for initializing the soft-in-soft-out algorithm of SW decoding.

Currently, research efforts are focused on the estimation of the Laplacian distribution parameter, considering that more accurate parameter estimation tends to improve the a priori probability estimation and hence reduce the bit rate required for the SW decoding. Among these works, the transform-domain adaptive correlation estimation (TRACE) algorithm in [14] has provided a relatively accurate way to estimate the parameter of the Laplacian correlation noise model through a refinement approach. In contrast to the CVC like H.264, which greatly benefits from the adoption of entropy approaching tools to exploit high-order statistical correlation among the coefficients such as using run-length coding for consecutive zeros, the correlation noise modeling in DVC can only exploit the first-order statistical correlation where only the co-located coefficient in the side-information frame is utilized to predict current coefficient, leaving high-order statistical correlation among coefficients being hardly exploited. Although some attempts like [114] have been made to exploit high-order inter-bit correlation by introducing a hidden Markov model at the bit-plane domain, the improvement is very limited in most cases, while the decoding complexity increases drastically.

In view of the importance of exploiting high-order statistical correlation among the coefficients as well as the limitation of existing DVC schemes, in this chapter we propose a novel approach to exploit high-order statistical correlation among the coefficients and their binary representations, thus achieving a better prediction of the bit a priori probability to increase the coding efficiency. More specifically, we first study the correlation modeling in wavelet domain DVC. By exploiting the self-similarity across the scales (bands), we can make better prediction of the coefficients of low magnitudes (insignificant coefficients) across scales. That is, in the wavelet transformed tree structure, insignificant coefficients are more likely to have insignificant children. Such inter-coefficient correlation is modeled and utilized for the binary symbol probability prediction. Meanwhile, when the DVC decodes from the least significant bit-plane to the most significant one, the previously decoded neighboring bits in less significant bit-planes can help to predict the insignificance of the coefficients especially in smooth regions and higher frequency bands. Such inter-bit correlation is also exploited to provide a more local prediction compared with the existing Laplacian correlation noise based prediction which relies on more global variance estimation. We finally integrate the

bit a priori probability estimation results of the three approaches, namely, the proposed inter-coefficient and inter-bit correlation exploitation scheme, as well as the widely used Laplacian correlation noise modeling. With the integrated modeling, the bit symbol a priori probability for initializing SW decoding can be estimated with a higher accuracy, leading to a better DSC efficiency. By implementing the proposed scheme to the DVC framework with wavelet domain multi-resolution motion refinement (MRMR) [24], significant coding gain of up to 14% bit rate reduction (equivalent to 1.2 dB gain) can be achieved compared with the original MRMR DVC scheme in [24].

The contributions of this chapter can be summarized as follows. First, we adapt the property of self-similarity among wavelet coefficients to DVC to effectively exploit the high-order inter-coefficient correlation, with the correlation modeling studied. Secondly, we consider the exploitation of high-order correlation among neighboring bits. By using the proposed correlation exploitation scheme, we demonstrate that the rate-distortion performance can be significantly improved.

The remainder of this chapter is organized as follows. In Section 4.2, we review the DVC framework with the state-of-the-art decoder motion estimation technique. In Section 4.3 we discuss the limitations of the existing Laplacian based correlation modeling and introduce the proposed correlation estimation scheme. Experimental results and related discussions are shown in Section 4.4. The summary of this chapter is presented in Section 4.5.

4.2 The Distributed Video Coding Framework

The framework used in this chapter is the same as the one described in Chapter 2. Note that in this framework, the group of picture (GOP) size can be arbitrarily large. Meanwhile, wavelet transform in [24, 71, 115]. is applied which is another popular choice for DVC apart from DCT. More importantly, we employ the recently developed decoder multi-resolution motion refinement (MRMR) algorithm [24] to estimate the side-information. With the MRMR based DVC system, the motion search has been totally shifted to the decoder, and a light weight encoding is achieved with good performance especially with high bit-rates. In the follow-

ing, we will provide more details on the key elements, namely, MRMR algorithm and the Laplacian distribution based correlation modeling.

4.2.1 Decoder Multi-Resolution Motion Refinement (MRMR) Algorithm

Figure 4.1: Decoder multi-resolution motion refinement (MRMR) algorithm.

In this section, we review the wavelet domain MRMR algorithm at the decoder side, which is illustrated in Fig. 4.1. Using the notations and steps presented in [24], the encoder applies an N -level wavelet decomposition to the WZ frames, yielding totally $(3N + 1)$ subbands for each of them. Besides the lowest frequency

subband LL_N , the high frequency bands are denoted as HL_n , LH_n , HH_n , ($n = 1, 2, \dots, N$) (reconstruction can be made from subbands LL_n , HL_n , LH_n and HH_n , ($n = 1, 2, \dots, N$) to produce the low frequency band LL_{n-1} in a lower decomposition level). These subbands are encoded and transmitted respectively using the WZ approach described earlier.

At the decoder side, with a reference frame $F(t-1)$ available, the current Wyner-Ziv $F(t)$ can be reconstructed as follows ((t) and $(t-1)$ denote the temporal indices). First, the reference $F(t-1)$ is decomposed using the same approach applied to $F(t)$. The decoding process starts from the lowest frequency band $LL_N(t)$. The side-information for band $LL_N(t)$ can be constructed using motion interpolation / extrapolation or directly using $LL_N(t-1)$. Since the lowest frequency bands from adjacent frames are generally highly correlated, limited coding efficiency loss can be expected even with coarsely estimated side-information. With the reconstructed $LL_N(t)$ by Wyner-Ziv decoding, the motion vectors are refined by motion estimation between $LL_n(t)$ and $LL_n(t-1)$, and the refined motion field is used to generate side-information for the three corresponding high frequency bands $HL_n(t)$, $LH_n(t)$ and $HH_n(t)$ at the same decomposition level. With better side-information quality, the bit rates for decoding these high frequency bands can be saved. With all the bands $LL_n(t)$, $HL_n(t)$, $LH_n(t)$ and $HH_n(t)$ decoded and reconstructed, $LL_{n-1}(t)$ at a lower decomposition level can be obtained using the inverse wavelet transform (not WZ decoding) and it is used to refine motion field for the three high frequency bands at level $n-1$ based on $LL_{n-1}(t-1)$. The process is repeatedly performed till all bands are decoded.

With the MRMR algorithm, the decoder can construct high quality side-information in a higher frequency band through motion refinement from the lower frequency band, based on the resolution scalability. Since the MRMR algorithm does not depend on the motion estimation between adjacent key frames and it well exploits the partial information of current frame generated during the decoding process, significant performance gain can be achieved over the interpolation / extrapolation based schemes. Owing to its superior performance and its consistency with our wavelet coefficients correlation exploitation structure, the MRMR based DVC framework is considered in this chapter.

4.2.2 The Laplacian Distribution Based Correlation Estimation

While better side-information quality which reduces the virtual channel errors, the coding efficiency of SW coder can be expected to increase. Furthermore, correlation modeling can further increase the coding efficiency which is analogous to providing information of the positions where these virtual channel errors occurs. In the context of practical DSC which is based on the channel coding techniques such as LDPC or turbo coding, the reliability of the side-information bits is used to initialize the decoding process. In particular, the reliability is represented by the log-likelihood ratio ($LLR = \log(p = 0/p = 1)$), where $p = 0$ and $p = 1$ are the estimated probabilities of a source bit to be 0 and 1, respectively. It has been shown that failure to provide LLR information accurately may significantly degrade the performance of the channel codes [116] which DSC is based on. In the existing DVC schemes, the estimation of LLR solely relies on the correlation modeling of correlation between the current WZ frame and its side-information frame. More specifically, the correlation between the original source X and its side information Y is commonly modeled through a Laplacian conditional density function:

$$f_{X|Y}(x|y) = \frac{\alpha}{2} e^{-\alpha|x-y|} \quad (4.1)$$

where the parameter α depends on the variance σ^2 of the correlation noise $Z = X - Y$:

$$\alpha = \sqrt{\frac{2}{\sigma^2}} \quad (4.2)$$

The estimation of σ^2 has been intensively studied in the sequence, frame and coefficient band domains [14] based on the side-information frame.

Furthermore, to achieve the minimum SW coding rate based on the Laplacian model, it is necessary to compensate for the coding efficiency loss when SW coding the binary representations of the Laplacian modeled decimal coefficients. Hence, inter-bit correlation exploitation within each coefficient is considered based on the Laplacian correlation modeling [2, 117]. The theoretical discussion of such inter-

bit correlation exploitation can be found in [118]. That is, consider a random variable X that can be represented by K bits “ $B_0(MSB), B_1, \dots, B_{K-1}(LSB)$ ”, where B_k are the random variables representing the binary digits in the K -bit representation of X (The same notations are used as in [118]). According to [118],

$$H(X) = H(B_{k-1}) + H(B_{K-2}|B_{k-1}) + \dots + H(B_0|B_1, \dots, B_{K-1}) \quad (4.3)$$

Hence, in DVC, with the side-information Y , we have

$$\begin{aligned} H(X|Y) = & \\ & H(B_{k-1}, Y) + H(B_{K-2}|B_{k-1}, Y) + \dots \\ & + H(B_0|B_1, \dots, B_{K-1}, Y) \end{aligned} \quad (4.4)$$

These equations suggest that the previously decoded bit-planes should be taken into account while decoding current bit-plane. They also suggest that there is no loss as long as such inter-bit correlation is fully exploited, regardless of the decoding-order (from MSB to LSB or in the reversed order). Note that the purpose of the existing inter-bit correlation exploitation in [2, 117] is to compensate the coding efficiency loss when coding decimal symbols using their binary representation, instead of providing additional information of current bits. Hence, context information from neighboring coefficients and bits is not actually considered by this existing inter-bit correlation exploitation scheme.

4.3 Proposed Correlation Estimation Scheme

As elaborated in the previous section, decoder side-information estimation and correlation modeling are the two key elements to the DVC performance. Unfortunately, existing schemes consider the previously decoded frames as the only side-information source, and model the correlation based on them. In this section, we will further analyze the limitations of such existing correlation estimation scheme and discuss the motivation for our high-order correlation exploitation scheme. We will then present our proposed correlation estimation scheme which can exploit

high-order inter-coefficient correlation effectively using more correlation modeling on more side-information sources.

4.3.1 Motivation of high-order correlation exploitation scheme

Since the existing Laplacian modeling only exploits the correlation between the current coefficient and the corresponding one in its side-information frame, it is actually a first-order correlation estimation scheme. This is more analogous to the exploitation of temporal correlation in the conventional CVC. However, the success of the CVC not only comes from the temporal correlation exploitation, but also significantly relies on the exploitation of correlation among the coefficients such as using the run-length coding and arithmetical coding. Therefore, one of the reasons for the performance gap between CVC and DVC is the fact that the Laplacian distribution based correlation model can hardly exploit the inter-coefficient correlation.

Inspired by the benefits of context based correlation exploitation in the conventional CVC, we propose a high-order correlation estimation scheme for DVC which effectively exploits the inter-coefficient correlation, instead of attempting to increase the accuracy of the Laplacian modeling as in most existing schemes. Our high-order inter-coefficient correlation estimation scheme comprising two approaches.

The first approach is to exploit the inter-coefficient correlation based on the self-similarity property. In the context of wavelet domain coding, the self-similarity of coefficients across scales can be taken advantage to exploit the high-order statistical correlation at the coefficient level by using the algorithms such as the zero-tree wavelet coding. More specifically, the self-similarity in wavelet domain can be described as: if a wavelet coefficient at a coarse scale (in low frequency band) is insignificant (with its magnitude less than a threshold), then all wavelet coefficients corresponding to the same orientation and spatial location at finer scales (in higher frequency bands) are likely to be insignificant (with their magnitude less than the threshold as well) [119]. That is, for the tree-structured representation of wavelet coefficients like in the zero-tree wavelet coding, the children are expected to have magnitudes not greater than their parent.

In conventional wavelet domain video coding, the self-similarity property has been efficiently exploited in the residual frames such as in [120].

On the other hand, in the DVC systems, such an important property has never been considered in the literature. As discussed earlier, if the self-similarity property can be used to provide the a priori probability estimation with more accuracy for the SW decoder, then higher coding efficiency of DVC is expected. Therefore, we propose to apply the self-similarity property in wavelet domain DVC. Our method starts from decoding the lowest frequency band using the MRMR algorithm. Based on the coefficients obtained in a lower frequency band and the self-similarity property, the corresponding children coefficients in a higher band can be predicted through a correlation modeling. The prediction for the current symbol with binary representation in the WZ frame can be obtained according to the inter-coefficient correlation modeling.

In the second approach, we consider predicting the current bit value using the context information of its neighboring bits (of neighboring coefficients) in the previously decoded bit-planes, which can have advantages in a smooth region. After obtaining the inter-coefficient correlation and inter-bit correlation based prediction, we integrate the observations from existing Laplacian modeling to obtain the final a priori probability prediction for the SW decoder initialization.

4.3.2 The Proposed Inter-coefficient Correlation Modeling

In this section, we present our inter-coefficient correlation modeling across scales in the wavelet domain, which aims to achieve a better prediction for probability of the magnitude of current symbol x_n given its parent's magnitude x_{n+1} in the lower frequency band (denoted as $p(x_n|x_{n+1})$). We first study the modeling of the distribution of $p(x_n|x_{n+1})$ based on some commonly used video sequences.

Three most popular distribution models are considered, namely, the Laplacian distribution given in (2.1), the zero mean Gaussian distribution ($p_{Gau}(x_n|x_{n+1})$) and the Cauchy distribution ($p_{Gau}(x_n|x_{n+1})$), as shown below:

$$p_{Gau}(x_n|x_{n+1}) = \frac{1}{\sqrt{2\pi}\sigma^2} e^{-\frac{x_i^2}{2\sigma^2}} \quad (4.5)$$

where σ^2 is the variance of x_n with a given parent x_{n+1} , and

$$p_{Cau}(x_n|x_{n+1}) = \frac{1}{\pi} \bullet \frac{\gamma}{x_n^2 + \gamma^2} \quad (4.6)$$

where γ is the scale parameter which specifies the half-width at half-maximum, and it can be obtained as $1/(\pi A_{max})$ given that A_{max} is the maximum amplitude of the distribution.

Quality of the distribution models is assessed by the Kullback-Leibler divergence (KLD), which is a measure criterion of the difference between two probability distributions P (the actual distribution) and Q (the approximated distribution). More specifically, KLD measures the expected number of extra bits required if the samples with distribution P are coded using a code optimal for distribution Q , rather than that for the actual distribution P . For probability distributions P and Q of a discrete random variable, KLD is defined as

$$D_{KL}(P||Q) = \sum_j P(j) \log_2 \frac{P(j)}{Q(j)} \quad (4.7)$$

The testing results using “foreman.qcif” sequence with a three level decomposition and a quantization stepsize of 10 are given in Fig. 4.2 and Table 4.1. To avoid clutter, only the distributions and approximations with parent magnitudes equal to 0 and 1 are shown in Fig. 4.2. From the figure and table, it can be seen that for the coefficients in Level 3 decomposition, Laplacian distribution provides a better approximation, while for the coefficients in Level 2 and Level 1 decomposition, Cauchy distribution provides a better approximation, which yields lower KLD values. It can also be seen that when the parent coefficient magnitude goes larger, it becomes more difficult to model the distribution of their children. Based the modeling, the binary symbol probability can be estimated using the same way as that for the Laplacian correlation noise model. The selection of distribution model and associated parameter can be estimated from the previously decoded frames by applying the KLD assessment online. We also find that performance is not quite sensitive to the accuracy of the estimation of distribution model parameters.

(a) Level 3 with parent=0

(b) Level 3 with parent=1

Figure 4.2: The distribution of the coefficients given their parent magnitudes.

(c) Level 2 with parent=0

(d) Level 2 with parent=1

Figure 4.2: The distribution of the coefficients given their parent magnitudes.

(e) Level 1 with parent=0

(f) Level 1 with parent=1

Figure 4.2: The distribution of the coefficients given their parent magnitudes.

Table 4.1: KLD for different distributions approximation

Level	parent	KLD(Lap)	KLD(Gau)	KLD(Cau)
3	0	0.50	0.87	0.61
	1	0.52	0.88	0.65
	2	0.51	0.88	0.63
	3	0.51	0.90	0.62
	4	0.56	0.95	0.66
	5	0.58	0.94	0.74
2	0	0.53	0.30	0.11
	1	0.56	0.36	0.11
	2	0.65	0.53	0.43
	3	0.75	0.68	0.64
	4	0.83	0.76	0.68
	5	0.88	0.87	0.75
1	0	0.43	0.41	0.09
	1	0.51	0.46	0.10
	2	0.56	0.53	0.33
	3	0.73	0.68	0.57
	4	0.87	0.89	0.76
	5	0.90	0.91	0.80

4.3.3 The Proposed Inter-bit Correlation Based Prediction Process

4.3.4 The proposed inter-bit correlation based prediction process

The existing inter-bit correlation exploitation approach [2, 117] described in the previous section considers only the bits belonging to the same coefficient, which aims to compensate for the loss introduced from converting coefficients in decimal representations to binary representations. Different from the existing approach, we propose to predict the current bit value considering its neighboring bits in the previously decoded bit-planes, where the bits belonging to the neighboring coefficients are also taken into account. A video frame may contain smooth regions which result in large amount of zeros in the quantized high frequency band

in the wavelet domain. The bits in less significant bit-planes may provide useful information to predict the bits in more significant bit-planes, which is desired to be exploited. In particular, if all the decoded less significant bits are zero, the more significant bit is more likely to be zero since the quantized high frequency coefficient tends to be zero instead of a large value. In such a way, neighboring information can be better exploited to provide a more local prediction compared with the existing Laplacian model based prediction which uses global information of the variances at bit-plane, frequency band and frame levels.

Therefore, we propose to exploit the inter-bit correlation for the prediction of zero coefficients. The proposed prediction process can be described as follows:

Figure 4.3: Context window.

(i) For coefficient $c(i, j)$ with bit-plane index i and location index j , define a context window W with size of $[i + 1, i + m] \times [j - n, j + n]$ as shown in Fig. 4.3, where $m - 1$ is the number of previously decoded bit-planes considered.

(ii) Based on the statistics obtained using the previously decoded frames, estimate the probability of a current bit being zero $p_0^i(k)$ given k ones observed in the context window W , where i is the bit-plane index.

(iii) Apply the prediction results to the bits in the current Wyner-Ziv frame.

Note that this prediction is made at the bit-plane level, which means bits in the same bit-plane have the same prediction given the identical context, while bits in different bit-planes may have different prediction results. In contrast to schemes like [14] which predict the a priori probability of current bit through the variance of a whole coefficient band or frame, our proposed inter-bit correlation exploitation scheme concentrates on local the neighboring information and predicts the a priori probability in a context based manner, which may provide better estimation accuracy.

4.3.5 Integrated Symbol A Priori Probability Prediction

Although the proposed inter-coefficient and inter-bit correlation based prediction can help to predict a priori probability for symbols especially in high frequency bands of smooth regions, there may still remain unpredicted or wrongly predicted bits (by the two proposed methods). Therefore, these two types of correlation should be integrated with the widely used Laplacian correlation noise modeling. In the context of DVC, let $P(x|y)$ be the estimated conditional probability of x given its side-information y , which is the well know Laplacian distribution based estimation, $P(x|c)$ be the estimation from the proposed inter-coefficient correlation based prediction, and $P(x|b)$ be the estimation from the proposed inter-bit correlation based prediction. (Note that more results from other types of estimations can be defined in the future). With the assumptions that these estimations results are independent (The existing Laplacian modeling is based on the side-information frame, while the two proposed modelings are respectively based on the parent coefficients and neighboring bits in the current WZ frame),

the final estimation can be obtained through Bayesian approach:

$$\begin{aligned}
p(0) &= P(x = 0|y, c, b) \\
&= \frac{P(y, c, b|x=0)P(x=0)}{P(y, c, b)} \\
&= \frac{P(x=0)P(y|x=0)P(c|x=0)P(b|x=0)}{P(y, c, b)} \\
&= \frac{P(x=0) \prod_{i \in \{y, c, b\}} P(x=0|i) \prod_{i \in \{y, c, b\}} P(i)}{P(y, c, b)P(x=0)^3} \\
&= \frac{\prod_{i \in \{y, c, b\}} P(i)}{P(y, c, b)P(x=0)^2} \prod_{i \in \{y, c, b\}} P(x = 0|i)
\end{aligned} \tag{4.8}$$

In this way, we can also obtain:

$$\begin{aligned}
p(1) &= P(x = 1|y, c, b) \\
&= \frac{\prod_{i \in \{y, c, b\}} P(i)}{P(y, c, b)P(x=1)^2} \prod_{i \in \{y, c, b\}} P(x = 1|i)
\end{aligned} \tag{4.9}$$

After the integration, the log-likely-hood ratio (LLR) which is used to initialize the SW (LDPCA) decoding process can be calculated as:

$$LLR = \log_2 \frac{p(0)}{p(1)} = \log_2 \frac{\prod_{i \in \{y, c, b\}} P(x = 0|i)P(x = 1)^2}{\prod_{i \in \{y, c, b\}} P(x = 1|i)P(x = 0)^2} \tag{4.10}$$

Assume $P(x = 0) = P(x = 1) = 0.5$, we finally obtain:

$$LLR = \log_2 \frac{\prod_{i \in \{y, c, b\}} P(x = 0|i)}{\prod_{i \in \{y, c, b\}} P(x = 1|i)} \tag{4.11}$$

4.4 Experimental Results

Preliminary experimental results are obtained (further tuning and checking are still on-going) to compare the proposed high-order inter-coefficient correlation exploiting scheme against the existing Laplacian distribution based correlation exploiting schemes as well as the conventional CVC standard H.264 in terms of the rate-distortion performance. Since the motion interpolation approach used in some prior works, e.g. in [111, 121], is incapable of processing video sequences

with large GOP size, the DVC schemes are implemented based on the MRMR algorithm.

Table 4.2: Prediction results and coding rates with example of the first fifteen coefficients in the frequency band HL_2 of the third frame of “foreman.qcif”

BP		c_1	c_2	c_3	c_4	c_5	c_6	c_7	c_8	c_9	c_{10}	c_{11}	c_{12}	c_{13}	c_{14}	c_{15}	rate
1	y	0	1	0	0	0	0	0	1	0	0	1	1	0	0	0	
	x	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	
	$p_x^y(0)$	0.998	0.016	0.998	1	1	1	1	1	1	1	1	0	1	1	1	0.303
	$p_x^x(0)$					0.968	0.968	0.940	0.940	0.997	0.997			0.997	0.997	0.997	
	$p_x^z(0)$				0.905	0.983	0.983	0.983	0.983	0.983	0.983				0.983	0.983	
	$p_x^w(0)$	0.998	0.016	0.998	1	1	1	1	1	1	1	1	0	1	1	1	0.303
2	y	0	1	0	0	0	0	0	1	0	0	1	1	0	0	0	
	x	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	
	$p_x^y(0)$	0.898	0.027	0.692	0.982	0.988	0.997	0.995	0.996	0.994	0.995	0.991	0.001	0.998	0.997	0.998	0.136
	$p_x^x(0)$					0.877	0.877	0.832	0.832	0.951	0.951			0.951	0.951	0.951	
	$p_x^z(0)$					0.954	0.954	0.954	0.954	0.954	0.954				0.954	0.954	
	$p_x^w(0)$	0.898	0.027	0.692	0.982	1	1	1	1	1	1	1	0.001	1	1	1	0.122
3	y	0	1	0	0	0	0	0	1	0	0	1	1	0	0	0	
	x	0	1	1	0	0	0	0	0	0	0	0	1	0	0	0	
	$p_x^y(0)$	0.513	0.263	0.758	0.774	0.766	0.973	0.942	0.223	0.931	0.896	0.224	0.178	0.902	0.925	0.937	0.621
	$p_x^x(0)$					0.813	0.813	0.721	0.721	0.907	0.907			0.907	0.907	0.907	
	$p_x^z(0)$																
	$p_x^w(0)$	0.513	0.263	0.758	0.774	0.934	0.994	0.977	0.426	0.993	0.988	0.224	0.178	0.989	0.992	0.993	0.533

The DVC systems based on MRMR motion estimation is implemented with the LDPCA codes [46] as SW coder, and the method described in section 4.2 is adopted for the Laplacian distribution based correlation estimation. Only the previous frame is used for motion estimation, and the search range is set to $[-16, 16] \times [-16, 16]$ with quarter-pixel motion search resolution. The same settings are applied to the H.264/AVC for comparison. Furthermore, for H.264, software JM13.2 is used with the baseline profile. The luminance component of several qcif sequences with varying motion activities are tested at 30 fps, with the first frame as key(I) frame and the all the subsequent frames as WZ (P) frames. Only the performance of the WZ (P) frames is shown in the following figures. We compare the proposed high order inter-coefficient correlation estimation scheme against the practical Laplacian modeling based on previously decoded frames (denoted as reference scheme 1) as well as H.264. The results are illustrated in Fig. 4.4. From the figure, it can be seen that at all rates, the proposed scheme outperforms the existing Laplacian correlation based scheme (reference scheme 1) significantly, which is on average 5 – 14% rate saving or around 0.5-1.2dB improvement in terms of distortion.

To provide more insights into the coding gain, we list the prediction of bit level symbol probability for decoding the coefficients in the band HL_2 of the 3rd

frame of “foreman.qcif” as in Table 4.2. Due to the space limitation, only the first 15 coefficients are shown. In Table 4.2. The probabilities of a symbol to be zero estimated using the existing Laplacian correlation modeling, the proposed inter-coefficient scheme, inter-bit scheme, and the hybrid scheme are denoted as $p_x^l(0)$, $p_x^c(0)$, $p_x^b(0)$ and $p_x^h(0)$, respectively. Note that the best or accurate estimation for $p_x(0)$ is 1 if x is zero and zero if x is 1. Since the bit-plane of sign can generally be correctly predicted, Table 4.2 only lists the bit-planes for representing the magnitudes of the coefficients, where bit-plane (BP) 1 denotes the most significant bit-plane and BP 1 denotes the most significant bit-plane.

From Table 4.2, it can be seen that the Laplacian correlation based probability estimation scheme can predict the most significant bit-planes quite well by using the knowledge from the less significant bit-planes. However, with the side-information as the only available knowledge, prediction is relatively poor for the least significant bit-planes, especially in some complex or high motion regions such as the prediction for c_1 to c_5 . Even in smooth or low motion regions such as the prediction for c_6 to c_{10} and c_{13} to c_{15} , the prediction efficiency is reduced. This results in high coding rates for the least significant bit-planes. On the other hand, with more information from the lower frequency bands as in the proposed scheme, more coefficients in the higher frequency bands are predictable can be predicted more accurately, such as c_9 , c_{10} and c_{13} to c_{15} , especially c_5 and c_8 in bit-plane 3. The inter-bit correlation estimation can also contribute to the prediction for some bits such as c_5 to c_{10} for bit-plane 2 to increase the accuracy. (Note that there is no contribution for bit-plane 3 because the inter-bit prediction requires previously decoded bit-plane.) Although the accuracy improvement is limited for a single bit, the contribution can be significant with large number of bits involved, which reduces the coding rates. With better predicted bits in the least significant bit-planes and the accuracy maintained for the most significant bit-planes, The reduced overall coding rate can be achieved.

The experimental results validate the effectiveness of the proposed scheme, which considers the correlation among the wavelet coefficients across scales as well as among the neighboring bits, with significant coding gain over existing DVC schemes. More importantly, it is demonstrated that high-order correlations should be exploited more efficiently to improve the performance of DVC systems.

4.5 Summary

Through our investigation, it has been found that the Laplacian correlation noise model, which was the only tool used in previous DVC schemes to exploit the statistical correlation, is insufficient to provide highly accurate symbol probability prediction to enhance the performance of DVC. By recognizing the importance of exploiting correlation using more context information, we have proposed a novel high-order correlation estimation scheme for wavelet domain DVC, which exploits the inter-coefficient correlation across scales inspired by the self-similarity property, as well as the inter-bit correlation based on the statistical information from the neighboring bits. All the correlation estimation results are further integrated to enhance the a priori probability prediction for the current bit which is used in the SW decoding, yielding higher coding efficiency. Our proposed scheme has been implemented on the MRMR based DVC framework. Our preliminary experimental results have demonstrated the success of the proposed scheme, by showing significant coding gain over the original MRMR scheme that employs the Laplacian correlation noise model only. With the results of our work, the future DVC research may focus on exploiting other types of high-order correlation in order to further narrow down the performance gap between DVC and CVC.

(a) “Foreman”

(b) “Mobile”

(c) “Stefan”

Figure 4.4: Rate-distortion performance comparisons for different sequences.

(d) “Carphone”

(e) “Coastguard”

(f) “Bus”

Figure 4.4: Rate-distortion performance comparisons for different sequences.

Chapter 5

A Joint Source-Channel Video Coding Scheme Based on Distributed Source Coding

In addition to the schemes of distributed video coding efficiency enhancement, in this chapter we consider one of the emerging applications of distributed video coding. By employing a single distributed video codec to compress the source and protect it against transmission errors, we present our joint source-channel coding scheme to tackle the drifting problem in video coding.

5.1 Introduction

The motion-compensated predictive coding principle has been widely used in modern video codecs to achieve high compression performance. However, it is also a vulnerable compression framework in the presence of transmission errors, because the correct decoding of any predictively coded symbols highly relies on the reconstruction quality of all reference symbols. One of the major problems posed in the practical communication of predictively coded signal over lossy channels is that of error propagation caused by predictive mismatch. More specifically, the mismatch between the reference symbol at the encoder and that at the decoder can lead to severe quality degradation of subsequently reconstructed data.

Figure 5.1: The error resilient video compression framework in [1].

Extensive efforts have been spent to tackle the above mentioned error propagation problem in encoding, transporting and decoding process [122]. The conventional schemes such as automatic-retransmission-request (ARQ) [123] and forward-error correction (FEC) [124] have their limitations. ARQ is not suitable for real-time scenarios since it suffers from a large delay due to feedbacks and retransmissions for lost or erroneous data. On the other hand, FEC is usually applied to the most important information such as header information and motion vectors due to a tight rate budget. Errors in the unprotected or less protected residual information will still cause the error propagation. Error concealment at the decoder side can also be employed to conceal the effect of errors to a certain degree. Recently, a new error resilient video coding technique has been received increasing research attention, which is based on distributed source coding (DSC) principles that independently generate the reference frame and the predicted frame to improve the robustness against the reference frame errors.

Apart from the low complex encoding studied in the previous chapters, another important application of DSC is on error resilient coding. A few robust video coding schemes using DSC techniques have been proposed to prevent error propagation in predictive video coding [1, 15] and protect enhancement layers in fine granularity scalability (FGS) coding [2, 16]. Furthermore, the link between predictive coding and distributed source coding has been elucidated in [1], which

considers the predicted symbol as a corrupted version of the reference symbol, and vice versa. An idea of forming a joint source-channel coding (JSCC) based on DSC technique has been proposed in [125, 126]. Better performance is expected with the JSCC scheme in practice although theoretically the source coding and channel coding can be optimized separately to achieve the same performance (with some impractical assumptions) as JSCC. However, in these existing error resilient video coding schemes, the SW codes are only applied after predictive compression. Therefore, the SW codes are mainly or totally employed as channel codes, while ignoring their source coding capability.

In view that the source coding can be done by considering the reference symbol as a corrupted version of the predicted (current) symbol through a virtual channel, while the source coding result has to be protected by channel codes when being transmitted through a real error-prone channel, it is therefore logical to consider fusing the virtual and real channels together to form a combined channel. Consequently, a single channel code can be applied to the combined channel to correct both the virtual and real channel errors [125]. In this chapter, such an error-resilient video compression scheme is developed that uses a unified SW codec for the JSCC. The SW codec not only exploits correlation between the current frame (Wyner-Ziv frame) and its reference frame (key frames), but also protects against channel errors. Based on the efficiency analysis of the JSCC scheme, the coding scheme is applied to certain selected frames in a video sequence to stop the error propagation more efficiently. These frames are intra-coded using SW codes with the knowledge of both the channel statistics and the correlation to their reference frames, which differs from the scheme in [1] where these selected frames are predictively encoded and DSC protected. The design of the channel-aware joint source-coding algorithm will be presented in details in the following sections.

The remainder of this chapter is organized as follows. In Section 5.2, we review and analyze some existing DSC-based error resilient video coding schemes. In Section 5.3 we analyze the DSC-based joint source-channel coding scheme. In Section 5.4, the details of coding design for joint source-channel video coding are presented. Experimental results are shown in Section 5.5. We summarize this chapter in Section 5.6.

5.2 Review of DSC-based Error Resilient Video Coding Schemes

In this section, two representative DSC-based error resilient video coding schemes are reviewed.

5.2.1 DSC-based Error Resilient Predictive Video Coding

In [1], a well known error resilient video compression framework using DSC technique was proposed to tackle the error propagation problem. This scheme is illustrated in Fig. 5.1. In this scheme, all the video frames are firstly encoded using the H.264 [127] video encoder. The first frame is intra-encoded as “I-frame”, and all the subsequent frames are encoded predictively as “P-frames”. Furthermore, the motion vectors and the header information for each frame are adequately protected so that they can be reconstructed correctly at the decoder. The scheme focuses on stopping error propagation caused by the erroneously decoded residual data in the video. Certain frames are selected as “peg” frames and they are used to stop the error propagation when video sequences are coded and transmitted through lossy channels.

While encoding non-peg video frames, the H.264 encoder is applied unalteredly. On the other hand, for each peg frame, besides the header information, motion vectors, and residual data generated by the standard H.264 video encoder, additional data known as coset information from SW coder is generated to stop error propagation. To generate such protection information, block-wise Discrete Cosine Transform (DCT) and dead-zone quantization are applied to the peg frame. The quantization result is converted to normal bit-plane, and each bit-plane is input to SW encoder (LDPC encoder) and the parity check bits generated are protected and transmitted. This procedure is analogous to the DSC schemes for light-weight encoding [15, 55, 128]. The differences are: 1) The error resilient scheme uses erroneously reconstructed peg frame as side-information, while the light weight encoding scheme uses previously reconstructed frame; 2) The error resilient scheme determines the rate of parity check bits based on channel conditions, while the light weight encoding scheme obtains the rate in the

light of the correlation between current frame and its side-information frame.

From above, it can be clearly seen that the compression is achieved by the conventional predictive coding approach, while the SW code is only used to protect the frame data, where source coding and channel coding are carried out separately.

5.2.2 DSC-based Error Resilient Fine Granularity Scalability Video Coding

Another well known DSC-based error resilient video coding scheme was proposed for the fine granularity scalability (FGS) coding [2] which is shown in Fig. 5.2.

Figure 5.2: DSC-based error resilient FGS video coding in [2].

In this scheme, the base layer is encoded by conventional H.264 video encoder, while the enhancement layer is encoded using the DSC technique in such a way that the enhancement layer can be reconstructed correctly even when the base layer is corrupted by channel errors. The compression of the enhancement layer is achieved mainly by the quantization and coset partition (binning) step. As to be discussed in Section 5.3, the coset partition can achieve the same coding efficiency as the fixed length predictive coding schemes if the correlation between the to-be coded source and its side-information is known, although generally predictive coding outperforms coset partition. Then the SW code (LDPC code) is applied to protect the indices of the cosets. As in [1], the rate of the LDPC code is mainly determined by the channel statistics. At the decoder, the reconstructed base layer (with or without errors) is used as the side-information for LDPC decoding to

reconstruct of the enhancement layer.

Although the DSC approach of coset partition is applied for data compression, its robustness against reference errors is quite limited if the errors at the reference are large, which makes the decoding error increase drastically. In our view, the error resilience of the scheme is mainly attributed from the LDPC code. Note that the LDPC code is not helpful in further compressing the coset indices since the indices correlation is very low. In this sense, the SW codes are applied separately for channel coding (LDPC code) and source coding (coset partition).

5.3 Analysis of the DSC-based Joint Source-Channel Coding Scheme

In view of the effectively separated source and channel coding in the existing DSC-based error resilient video coding schemes discussed above, we propose a joint source-channel video coding scheme where a single SW code is applied to perform both source coding and channel coding simultaneously. In particular, the reference can be considered as a corrupted version of the current information and a virtual channel can be formed. Considering the reference may be corrupted by real channel errors, the virtual channel and the real channel can be fused to make a combined channel. A JSCC can be applied to the combined channel [125].

In the following, we analyze the effectiveness and achievable efficiency improvement of the joint source-channel scheme, while we will give the design for video coding in Section IV.

5.3.1 Efficiency Analysis of DSC-based Joint Source-Channel Coding vs. Predictive Coding with FEC in the Case of Channel Errors at the Reference Only

We consider the coset partition [107, 108] to illustrate qualitatively the effectiveness of DSC-based JSCC scheme. Assume that the source symbol $X \in \mathbb{Z}$ is a quantized observation at the encoder with its reference symbol (side-information) Y only accessible at the decoder, and X and Y are correlated with $|X - Y| < d$.

To encode X by coset partition which is an interleaved quantizer, the coset index is obtained by

$$I = X \text{Mod}(D) \quad (5.1)$$

Hence, the coset index represents values with a minimum distance of D from each other, which is transmitted instead of the difference $X - Y$ as conventional approaches. As long as $d < D/2$, we can reconstruct X error-free from the coset indicated by the received index by using the value closet to the reference Y . An example of the coset partition is illustrated in Fig. 5.3a, where the top line shows the set of possible values for X using small circles, and the following lines show cosets which partition the set. The small solid box shows the actual value, and the actual value lies in the first coset in the example. The dashed box shows the decoded value. This coset partition method requires a minimum number of $\lceil \log_2(D) \rceil$ bits to encode X which is the same as that needed for a given reference Y of the conventional predictive coding based on the fixed length code. In this sense, there is no compression efficiency loss in the absence of noise for the coset partition with a known correlation between the source and its reference.

Now we consider the case that the received reference is corrupted by real channel noise n with $|n| < N$. In the conventional predictive coding, the noise will lead to mismatch between the encoder and the decoder. Reconstruction errors will propagate to all the subsequent symbols dependent on this reference. For the predicted symbol X to be reconstructed correctly, channel codes must be applied to the reference to correct the noise. In contrast, using coset partition, the reconstruction error can be eliminated by simply increasing the minimum distance D . From Fig. 5.3b in the case of noise n corrupting reference Y , it can be seen that as long as $d + N < D/2$ (corresponding to the shadowed region with length D), the noise can be corrected. D can be increased with more cosets and hence more bits are used to represent the coset indices.

The example above has illustrated qualitatively the robustness against errors at reference using the DSC approach. Now we will show the improvement of coding efficiency quantitatively. Consider correlated binary sequences $X = x_1x_2x_3\dots x_L$ and $Y = y_1y_2y_3\dots y_L$ with i.i.d. and memoryless random variables and block length L . Compression can be achieved by coding X dependent

(a)

(a) Noise free

(b)

(b) Noise at the reference

Figure 5.3: An example of error resilient compression using coset partition.

on Y . Theoretically, the minimum rate required to make X decodable given reference Y at the decoder is the conditional entropy $H(X|Y)$ as a function of the conditional probability p obtained as:

$$p = D_H/L \quad (5.2)$$

where D_H is the Hamming distance between X and Y . Hence,

$$H(X|Y) = H(p) = -p \times \log(p) - (1 - p) \times \log(1 - p) \quad (5.3)$$

It is known in DSC, a virtual Binary Symmetric Channel (BSC) with crossover probability p_v is assumed to connect X and its side-information Y . $H(p_v)$ is also the Slepian-Wolf bound [39]. Note that for any entropy function $H(p)$ discussed throughout this chapter, we always consider $0 < p < 0.5$.

The compressed X of length $H(p_v) \times L$ and its reference Y of length L are sent. We consider the real channel errors occur at the reference Y only without affecting the difference $X - Y$. In the case of a real BSC with crossover probability p_r , additional channel codes must be applied to Y in order to combat channel errors in conventional predictive coding. To correctly reconstruct the predicted symbol X , protected Y and compressed X need to be sent at the total bit rate

$$R_{pred} = \frac{L}{C} + L \times H(p_v) = L \times \left(\frac{1}{1 - H(p_r)} + H(p_v) \right) \quad (5.4)$$

where the BSC capacity $C = 1 - H(p_r)$, and L/C is the bit rate for protection while $L \times H(p_v)$ is the rate for compressing X .

The virtual BSC with crossover probability p_v and the real BSC with crossover probability p_r can be fused to form a combined channel with the crossover probability being $p_v + p_r - 2p_v p_r$, when the virtual channel errors and the real channel errors are independent. With the same assumption as above that the real channel errors affect the reference only without distorting the syndrome or parity check bits generated by an SW encoder, the total number of bits required to correctly reconstruct the coded symbol X is

$$R_{joint} = L + L \times H(p_v + p_r - 2p_v p_r) = L \times [1 + H(p_v + p_r - 2p_v p_r)] \quad (5.5)$$

Note that the reference Y is not protected which may be corrupted by channel noises. We know that $1/[1 - H(p_r)] > 1 + H(p_r)$ for $0 < H(p_r) < 1$, which leads to $R_{pred} > L \times [1 + H(p_r) + H(p_v)]$. The entropy function $H(p)$ is concave and monotonically increasing for $0 < p < 0.5$ [105], which leads to $H(p_v + p_r - 2p_v p_r) < H(p_v + p_r) < H(p_v) + H(p_r)$. Therefore, we can obtain $R_{joint} < L \times [1 + H(p_r) + H(p_v)] < R_{pred}$, which suggests the total rate required to correct the reconstruction error of the predicted symbol X for terminating the error propagation can be reduced with the DSC-based joint source channel coding

scheme.

5.3.2 Comparative Efficiency Analysis in a More General Case

In a more general case, where the channel errors may corrupt both the reference and the syndrome (parity check) bits generated by an SW encoder, we consider a series of $M+1$ correlated binary information sequences $X^{(m)} = x_1^{(m)} x_2^{(m)} x_3^{(m)} \dots x_L^{(m)}$ with block length L , where $m \in [0, M]$. Dependent compression can be applied to $X^{(m)}$ based on its previous sequence $X^{(m-1)}$. The virtual channel crossover probability between $X^{(m)}$ and its reference $X^{(m-1)}$ is denoted as $p_v^{(m)}$. The minimum rate required to make $X^{(m)}$ decodable can be obtained using (5.2) and (5.3) given $X^{(m-1)}$ at the encoder.

Now we compare the rates required for different approaches to transmit the $M+1$ sequences $X^{(0)}, X^{(1)}, X^{(2)}, \dots, X^{(M)}$ and reconstruct $X^{(M)}$ correctly for stopping error propagation. The first sequence $X^{(0)}$ is not coded, and the following sequences $X^{(m)}$ are coded dependently on their previous sequences $X^{(m-1)}$.

For the predictive coding scheme where the sequence $X^{(M)}$ is independently coded (i.e., intra refreshed) and protected to stop error propagation, leaving the previous symbols unprotected, the number of bits required is

$$R_{intra.rf} = L \times \left(1 + \sum_{m=1}^{M-1} H(p_v^{(m)}) + \frac{1}{1 - H(p_r)} \right) \quad (5.6)$$

On the other hand, according to the scheme in [1], where the inter-coding of $X^{(M)}$ is performed based on $X^{(M-1)}$ while the SW code is generated to protect $X^{(M)}$, the number of bits required is

$$R_{[1]} = L \times \left(1 + \sum_{m=1}^M H(p_v^{(m)}) + \frac{H((M+1)p_r)}{1 - H(p_r)} \right) \quad (5.7)$$

Note that the previous sequences are not protected as that for the intra-refresh scheme and the syndrome (parity check) bits generated by the SW encoder are protected by channel codes. The accumulated real BSC after M times of error

propagation (with $M + 1$ erroneous sequences) is approximated as a cascade of $M + 1$ BSCs, each having crossover probability p_r (the real channel crossover probability is used to approximate the error rate of reconstructed predictive differences). Thereby, the accumulated real BSC has a crossover probability of $\frac{1}{2}(1 - (1 - 2p_r)^{M+1}) = (M + 1)p_r + \sum_{k=2}^{M+1} f_k(M + 1)p_r^k$, where $f_k(M + 1)$ is the coefficient of p_r^k . Since small p_r values (less than 0.1) are considered in this chapter, which lead to $p_r \gg p_r^k$ for $k \geq 2$, we further approximate the accumulated real BSC crossover probability to $(M + 1)p_r$.

Now we consider the DSC-based joint source channel coding approach, where the virtual BSC with crossover probability $p_v^{(M)}$ and the real BSC with crossover probability p_r are fused to form a combined channel with the crossover probability $p_v^{(M)} + Mp_r - 2p_v^{(M)}Mp_r$. The accumulated real BSC crossover probability is approximated as Mp_r since the side-information sequence for the SW coder is obtained after $M - 1$ times of error propagation. The total number of bits required to correctly reconstruct the predicted sequence $X^{(M)}$ is

$$R_{joint} = \tag{5.8}$$

$$L \times \left(1 + \sum_{m=1}^{M-1} H(p_v^{(m)}) + \frac{H(p_v^{(M)} + Mp_r - 2p_v^{(M)}Mp_r)}{1 - H(p_r)} \right)$$

It has been shown experimentally in [1] that the scheme in [1] is more efficient than the intra refresh scheme for stopping error propagation in terms of bit rate, that is $R_{[1]} < R_{intra_rf}$ for highly correlated symbols such as frames in a video sequence. To compare the rate for our proposed scheme R_{joint} against that for the scheme in [1] $R_{[1]}$, we carried out a Matlab simulation with varying variables, and observed that the relationship $R_{joint} < R_{[1]}$ always holds without any counter-example being found, although we have not found a mathematical proof for that. For an illustrative comparison, we plot the rates as a function of real channel crossover probability p_r for different M and $p_v^{(M)}$, assuming $L = 1$ and $p_v^{(m)} = p$ where $m \in [1, M]$. Since the two schemes are using the same method to code the non-peg frames, the plots only show the rates for peg-frames, which are $R_{[1]}^{peg} = L \times \left(H(p_v^{(M)}) + \frac{H((M+1)p_r)}{1 - H(p_r)} \right)$ and $R_{joint}^{peg} = L \times \frac{H(p_v^{(M)} + Mp_r - 2p_v^{(M)}Mp_r)}{1 - H(p_r)}$. From Fig. 5.4, it can be seen that our proposed scheme is more efficient than the scheme

in [1], and the rate improvement becomes more visible as p_r increases. We have also tested with many other different settings of M , p and p_r , which exhibit the similar comparison results (not shown in the figure to avoid clutter).

5.4 DSC-based Joint Source Channel Coding Design for Video

As analyzed above, a single SW code is able to accomplish both the data compression (source coding) and protection (channel coding) tasks jointly in a more efficient way. We now apply the DSC-based joint source-channel coding scheme for error resilient video coding and present the codec design in details.

As in [1], we consider the channel aware scenario where the real channel statistics is known, and in practise it can be obtained using training sequences. Same as that in [1], we use certain P-frames known as “peg” frames to stop the error propagation, and we assume that the first frame is an I-frame with all the subsequent non-peg P-frames $F^{(m)}$ inter-coded predictively from previous frames $F^{(m-1)}$, where m is the frame number. Different from [1], where the peg frames are also inter-coded with protection by DSC scheme, the peg frames in our proposed scheme are intra-encoded using a unified SW code with the knowledge of the channel statistics and the source correlation. By assuming that the motion vectors and the header information for each frame are adequately protected so that they can be reconstructed correctly at the decoder as in [1], we focus on correcting the reconstruction errors caused by the loss of residual data during the transmission.

The LDPC code with belief propagation decoding algorithm has been shown to be a near Shannon Capacity error correcting code, and widely applied in distributed source coding with performance close to the SW bound. A recently proposed LDPC-Accumulated (LDPCA) code [46] which accumulates the syndrome bits from conventional LDPC code outperforms the LDPC code by a small margin with the advantage that different rates can be achieved without altering the generating matrix. We will consider the LDPCA code as the SW code in the following section.

(a)

(a) $M = 1$

(b)

(b) $M = 4$

Figure 5.4: Comparisons of $R_{[1]}^{peg}$ and R_{joint}^{peg} .

Figure 5.5: The proposed encoder for DSC-based joint source-channel video coding.

5.4.1 Encoder

The encoder of the proposed scheme is shown in Fig. 5.5. For non-peg frames, H.264 (baseline profile) video encoder are applied. For peg frames, our scheme only transmits syndrome bits generated by a single SW encoder (LDPCA encoder) without sending any residual data from H.264 encoder. The encoder of the peg frames contains the following four major constituents.

5.4.1.1 Motion Estimation and Prediction

Motion estimation and prediction based on H.264 is performed and motion predicted frame F_{pred} is generated to obtain a good reference for the DSC coding of the peg frames. Furthermore, the motion vectors for the peg frames are adequately protected (thus no error) and transmitted as that for non-peg frames. Note that we are emphasizing on the error resilient video coding, the peg frame is allowed to communicate with its reference unlike the situation in DSC systems with light weight encoder.

5.4.1.2 DCT and Quantization

4×4 block-wise DCT is performed to both the peg frame F_{peg} and motion predicted frame F_{pred} , and the DCT coefficients are quantized using the H.264 dead-zone quantizer.

5.4.1.3 Forming Bit-Planes

After block-wise DCT and quantization, the peg frame consists of K 4×4 DCT blocks, and each block comprises 16 coefficients. The DCT coefficients with the same coordinates from each block are grouped and form vectors with length K . Each element of a coefficient vector is converted to its binary representation and bit-planes with length K are formed. Bit-planes are generated for the motion predicted frame using the same method.

5.4.1.4 Joint Source-Channel Encoding with LDPCA Code

Before we apply the LDPCA code to encode the bit-planes, the number of parity check bits needs to be determined. We assume the real Binary Symmetric Channel (BSC) crossover probability to be p_r . Note that other types of channels can be converted to BSC. For example, the conversion between BSC and AWGN channel can be made based on the equation:

$$\left(\frac{E_s}{N_0}\right) = -\log(2\sqrt{p_r(1-p_r)}) \quad (5.9)$$

where $\frac{E_s}{N_0}$ is the Symbol Error Ratio of the AWGN channel. If one frame is selected to be a peg frame for every $M + 1$ frames, $M \times p_r$ will be approximate as the accumulated real BSC crossover probability with small p_r used.

(5.2) can be applied to obtain the virtual BSC crossover probability $p_v^{i,j}(l)$ between the $l - th$ corresponding bit-planes of DCT coefficient with coordinates i, j from peg frame and its motion predicted reference frame. Total crossover probability $p_{total}^{i,j}(l)$ of the combined channel can be obtained as $p_v^{i,j}(l) + M \times p_r - 2p_v^{i,j}(l) \times Mp_r$. The rate of the LDPCA code can be obtained by applying (5.3) to $p_{total}^{i,j}(l)$. If $p_{total}^{i,j}(l)$ is greater than a threshold, it means the correlation is too low so that the bit-plane of the peg-frame could not be coded with DSC

approaches. In this case, the bit-plane will be intra-coded and protected instead if it is related to a significant bit-plane of a low frequency coefficient, or it will be discarded otherwise. In our experiments, the value of Th was set to be 0.25. which is found to have little effect on the results as long as the selection of the threshold is in the range of 0.23-0.25.

Different from the scheme in [1], where the SW code is used to be a channel code to protect the peg frames as auxiliary to the residual data inter-encoded by H.264 video codec, our scheme compresses and protects the peg frames at one go based on DSC without the inter encoding. In other words, our scheme uses a unified SW code to achieve joint source-channel coding.

5.4.2 Decoder

For non-peg frames, the standard H.264 video decoder is applied. The structure of the proposed peg frame decoder is shown in Fig. 5.6.

Figure 5.6: The proposed decoder for the peg frames.

The decoder first generates the motion predicted reference frame using previously reconstructed frame and the received motion vectors of the peg frame. Next, Block-wise DCT, quantization, bit-plane organization are performed in the same way as that at the encoder. With the inputs of the bit-planes and the received syndrome bits of the peg frame, LDPCA decoder generates the bit-planes of the peg frame. Finally, the reconstruction step includes converting bit-planes of peg frame to DCT coefficients, de-quantization, inverse DCT, and followed by optional de-blocking filtering as that in H.264.

5.5 Simulation Results

Simulations are carried out to compare the proposed joint source-channel video coding schemes against the DSC-based scheme [1] for eliminating the error propagation problem. Simulation results are provided for six different qcif sequences encoded at 30fps. One frame is set to be a peg frame for every five frames, e.g., Frame 5,10,15,..., which are used to recover from error propagation.

The coded video stream is transmitted over a BSC with crossover probability $p_r = 0.01$. Fig. 5.7a and Fig. 5.7b plot the decoded in peak signal-to-noise ratio (PSNR) results of Frame 1 to 30 for “Foreman” and “Mother and daughter” sequences, respectively. As can be seen from the figure, the peg frames are successfully recovered at Frame 5, 10, 15, ..., 30 for both the scheme in [1] and our proposed method, thus stopping the error propagation. The negligible loss in PSNR for the recovered peg frames is due to discarding some low correlated bit-planes. We use the LDPCA codes proposed in [46] for all the simulations. The rate of LDPCA code is calculated as discussed the previous section. A small amount of feedbacks for addition syndrome bits of the LDPCA code are allowed to achieve the successful LDPCA decoding. We also use Reed-Solomon code to protect the syndrome bits and the parameters of the code are adjusted to ensure the syndrome bits are received correctly.

Fig. 5.8 compares the number of bits required for the peg frames to recover from error propagation with the two compared schemes for $p_r = 0.01$. From the plots, it can be seen that our scheme saves up to about 10% bits than the scheme in [1] which shows our scheme is more efficient. Note that bits used for non-peg frames are the same for the two approaches since the same coding method of H.264 coder is applied. Compared to our analytical simulation results plotted in Fig. 5.4b, which shows about 20% rate improvement at $p_r = 0.01$, the rate improvement in the practical video coding is reduced, of which the major reason is that the relatively poor coding efficiency of the SW coding (LDPCA coding) compared to that of the entropy coding used in H.264 discounts the proposed joint source-channel coding gain.

Fig. 5.9 presents the rate-distortion performance comparison of peg frames for different sequences. The rate-distortion performance with the conventional FEC

(a)

(a) "Foreman"

(b)

(b) "Mother and daughter"

Figure 5.7: Decoded PSNR comparisons with $QP = 28$ and real BSC crossover probability $p_r = 0.01$.

(a)

(a) "Foreman"

(b)

(b) "Mother and daughter"

Figure 5.8: Peg frame bit rate comparisons with $QP = 28$ and $p_r = 0.01$.

code protected intra refresh approach is also in Fig. 5.9. For a fair comparison,

the peg frame rate and intra refresh rate are the same as $6fps$ for the $30fps$ video sequences in view of one peg frame among every five frames. In Fig. 5.9, the QPs are 16, 20, 24, 28 for “Foreman”, “Mother and daughter”, “Hall monitor” sequences and 24, 28, 32, 36 for “Coast guard”, “Stefan”, “News” sequences, respectively. It can be seen that in the case of channel errors, our scheme consistently outperforms the scheme in [1] due to less rate required to recover from channel errors. The improvement tends to be more significant with higher rates.

5.6 Summary

Recently, several error resilient schemes have been proposed to tackle the error propagation problem in the motion-compensated predictive video coding based on a promising technique – distributed source coding (DSC). However, these schemes mainly apply the distributed source codes for channel error correction, while under-utilizing their capability for data compression. A channel-aware joint source-channel video coding scheme based on DSC is proposed to eliminate the error propagation problem in predictive video coding in a more efficient way. It is known that near Slepian-Wolf bound DSC is achieved using powerful channel codes, assuming the source and its reference (also known as side-information) are connected by a virtual error-prone channel. In the proposed scheme, the virtual and real error-prone channels are fused so that a unified single channel code is applied to encode the current frame thus accomplishing a joint source-channel coding. Our analysis of the rate efficiency in recovering error propagation shows that the joint scheme can achieve a lower rate compared with performing source and channel coding separately. Simulation results show that the number of bits used for recovery from error propagation can be reduced by up to 10% using the proposed scheme compared to Sehgal-Jagmohan-Ahuja’s DSC-based error resilient scheme.

(a) "Foreman"

(b) "Mother and daughter"

(c) "Hall monitor"

Figure 5.9: Peg frame rate-distortion comparisons with different schemes ($p_r = 0.01$).

(d)

(d) "Coast guard"

(e)

(e) "Stefan"

(f)

(f) "News"

Figure 5.9: Peg frame rate-distortion comparisons with different schemes ($p_r = 0.01$).

Chapter 6

Conclusions and Future Works

6.1 Conclusions

Motivated by some emerging applications which require encoder with low complexity and power consumption, distributed source coding which employs a reversed architecture compared with conventional centralized source coding, has received increasing research interests. Inspired by the theoretical results and the popularity of portable wireless communications, the application of DSC on video, namely, distributed video coding (DVC) has become a promising paradigm to investigate encoders with low complexity and high coding efficiency for wireless video communications. Meanwhile, DVC can find more applications in the areas of error resilient video coding, scalable video coding, multiple description video decoding and multi-view video coding.

Despite the encouraging DSC theories and its successful coding results for ideal cases, there is still a remarkable performance gap between DVC and traditional centralized video coding (CVC). This performance difference is actually caused by the channel-code based distributed source coding inefficiency, side-information estimation unreliability and correlation prediction inaccuracy, which should be addressed to enhance the performance of the DVC systems.

In this thesis, we have studied three schemes to enhance the coding efficiency of distributed video coding by using the adaptive nested lattice quantization as well as exploiting high-order statistical correlation. We have also studied the

error resilient video coding by applying distributed source coding based joint source-channel coding approach. Major contributions are summarized as follows.

Firstly, by analyzing the two different coset partition methods, We have proposed an adaptively integrated coset partition approach for distributed video coding based on the prediction of block-wise correlation noise in video sequences. The proposed adaptive scheme attempts to maintain the advantages of both the schemes, while suppressing their respective disadvantages. In particular, a large decoding error caused by coset channel coding due to a large correlation noise can be avoided by switching to the SQ based binning method, while the conventional coset channel coding can achieve binning-error-free decoding with high compression ratio in the case of a small correlation noise. The encoder and decoder design based on the proposed scheme has been presented. In addition, by further studying and analyzing the two binning approaches, an adaptive 1-D nested lattice quantization scheme has been proposed according to the Laplacian distribution model of correlation noise between the source and its side-information. The optimized selection of parameters in the proposed scheme has been developed using a coarse approximation model for the correlation noise. The experimental results have demonstrated that the proposed adaptive schemes are able to achieve significant coding gain over the non-adaptive DVC schemes.

Secondly, through our investigation, we have found that the Laplacian correlation noise estimation is insufficient to provide highly accurate symbol probability prediction to enhance the performance of DVC, which was previously believed to be the only key to exploit the statistical correlation. By recognizing the importance of exploiting high order correlation, we have proposed a novel high order correlation estimation scheme in wavelet domain DVC, which exploits the inter-coefficient correlation across scales inspired by the self-similarity property, as well as the inter-bit correlation based on the statistical information from the neighboring bits. All the correlation estimation results are further integrated to obtain the enhanced a priori probability prediction for current bit which is used in the SW decoding for higher coding efficiency. Our proposed scheme has been implemented on the MRMR based DVC framework. Experimental results have justified the success of the proposed scheme, by showing significant coding gain over the original MRMR scheme. With the results of our work, the future DVC

research may focus on the exploiting other types of high order correlation in addition to the modeling of Laplacian distribution in order to further narrow down the performance gap.

Thirdly, by using the distributed source coding technique, where a virtual channel is assumed to connect the current information and its reference, a combined channel can be fused with the virtual channel and the real channel. Therefore, a unified single channel code may be applied to this combined channel for joint source-channel coding. Based on this principle, we have proposed a channel-aware error resilient video compression scheme using DSC-based technique. By applying a single SW code to peg frames for both compression and data protection purposes, these frames are used to stop error propagation. We have analyzed the rate efficiency improvement of the joint source-channel coding scheme over the separate source and channel coding scheme. Simulation results have validated the proposed scheme which requires a lower bit rate to successfully recover from error propagation compared to Sehgal-Jagmohan-Ahuja's scheme in [1] in the event of channel errors.

6.2 Recommendations for Future Work

Some possible directions of the future DVC research are listed as follows:

6.2.1 Side-information Estimation

The long standing and key problem for DVC is the side-information (SI) estimation, and the difficulty lies in that decoder must construct SI from previously decoded data with minimum distance to the Wyner-Ziv (WZ) encoded data which it does not know. The recent report has demonstrated the progressive motion refinement scheme [24], which learns motion for higher layers from the lower layers, is more promising than other widely used approaches such as CRC based, hash based and interpolation/extrapolation based ones. More accurate decoder motion estimation results are expected by improving such approach, including enhancing the initial SI construction quality for the lowest frequency band, developing advanced learning method to obtain better motion prediction for higher

bands based on lower bands. Meanwhile, since no motion vector is required for transmission in DVC, multiple references are possible. Hence, taking advantage of multiple references is another important way to improve SI quality. In addition, more motion estimation modes can be employed in DVC than that in CVC, such as varying block sizes, rotating and zooming modes.

6.2.2 Correlation Estimation

As described earlier, an accurate knowledge of the correlation between source and side information is essential to the coding efficiency of the channel coding based SW coder, to minimize reconstruction errors during de-quantization using side-information, and to achieve encoder rate control of the SW code. Therefore, the rate-distortion (RD) performance of a DVC system strongly depends on its capability to estimate the correlation model and its parameters. Our research has shown the significance of exploiting high-order correlation among the sources. Apart from the wavelet domain inter-coefficient and inter-bit correlation exploited in our work, more types of correlation are expected to be exploited not only in wavelet domain, but also in DCT domain and pixel domain.

6.2.3 Rate-distortion Optimization for DVC

It is well known that the traditional video coding benefits a lot from its rate-distortion optimization (RDO). However, RDO has never been considered in DVC. To achieve a RDO framework for DVC, adaptive mode design and selection with high coding efficiency and low complexity should be considered. Different modes in block level, frequency band level and bit-plane level for adaptive coding strategies may be designed taking into account the varying characteristics of video signal especially for high motion and occlusion regions.

6.2.4 Error Resilient Video Coding

Recently, distributed source coding based error resilient video coding has attracted a lot of research interests with approaches such as joint source-channel schemes and joint multiple description-distributed video coding schemes. Al-

though in these schemes, encoders are allowed to communicate with each other, feedback channel is still utilized to achieve the best performance. However, the feedback channel may be unavailable in some practical situations. Therefore, removing the feedback channel should be considered for these cases. More specifically, encoder need to estimate the required rate taking into account the channel conditions, the quality of side-information and the possible overhead information to achieve a rate-distortion optimized performance.

6.2.5 Removal of feedback channels in SW coding

Currently, good performance of DVC/DSC systems highly relies on the usage of feedback channels, which ensure successful decoding with minimal coding rates. However, transmission of addition bits upon requests from the feedback channel and re-decoding can cause large delay, and more importantly, the feedback channel may not available in many practical scenarios. Therefore, a future work could be removal of the feedback channel at limited cost of rate-distortion performance loss. Apart from the correlation noise based encoder rate estimation techniques described in Chapter 2, it is necessary to investigate rate-distortion performance of the SW codes in the case of unsuccessful decoding. In particular, SW codes should be designed to minimize the decoding error when insufficient protection bits are received.

Author's Publications

Journal Paper

1. Y. Zhang and C. Zhu, "A Joint Source-Channel Video Coding Scheme Based on Distributed Source Coding," *IEEE Trans. Multimedia*, vol. 10, no. 8, pp. 1648-1656, Dec. 2008.
2. Y. Zhang and C. Zhu, "Adaptive Coset Partition for Distributed Video Coding," *Signal Processing*, Vol. 90, Issue 8, pp. 2480-2486, Aug. 2010.
3. Y. Zhang and C. Zhu, "Adaptive One-Dimensional Nested Lattice Quantization for Distributed Video Coding," submitted to *Signal Processing: Image Communication*, under major revision.
4. Y. Zhang, C. Zhu and W. Zeng "Exploiting High-Order Correlation in Distributed Video Coding," submitted to *IEEE Trans. Circuits Syst. Video Technol.*, under major revision.

Conference Paper

1. Y. Zhang and C. Zhu, "Distributed video coding based on adaptive binning," *IEEE International Symposium on Circuits and Systems*, Kos, Greece, May 2006.
2. Y. Zhang and C. Zhu, "Full Search of Side-Information in Distributed Video Coding," *International Conference on Image and Graphics*, Chengdu, China, Aug. 2007.
3. Y. Zhang and C. Zhu, "Quantizer design for correlation noise in distributed video coding," *IEEE International Symposium on Broadband Multimedia Systems and Broadcasting*, Germany, 2011.
4. Y. Zhang, C. Zhu and Y. Xu, "Multiple Description Wyner-Ziv Video

Coding Based on Decoder Multi-Resolution Motion Refinement,” *IEEE International Symposium on Broadband Multimedia Systems and Broadcasting 2012*, Seoul, Korea, Jun. 2012.

Bibliography

- [1] A. Sehgal, A. Jagmohan, and N. Ahuja, “Wyner-Ziv coding of video: An error-resilient compression framework,” *IEEE Trans. Multimedia*, vol. 6, no. 2, pp. 249–258, Feb. 2004. [vii](#), [4](#), [25](#), [55](#), [88](#), [89](#), [90](#), [91](#), [96](#), [97](#), [98](#), [102](#), [103](#), [106](#), [111](#)
- [2] Z. X. Xiong and Q. Xu, “Layered Wyner-Ziv video coding,” *IEEE Trans. Image Processing*, vol. 15, no. 12, pp. 3791–3803, Dec. 2006. [vii](#), [4](#), [5](#), [23](#), [29](#), [35](#), [41](#), [43](#), [47](#), [51](#), [55](#), [70](#), [71](#), [78](#), [88](#), [91](#)
- [3] X. Artigas, J. Ascenso, M. Dalai, S. Klomp, D. Kubasov, and M. Ouaret, “The DISCOVER codec: Architecture, techniques and evaluation,” in *Proc. Picture Coding Symposium*, Lisbon, Portugal, Nov. 2007. [viii](#), [14](#)
- [4] P. L. Dragotti and M. Gastpar, *Distributed source coding: Theory, algorithms, and applications*. Academic Press, 2009. [viii](#), [1](#), [14](#), [23](#)
- [5] D. Slepian and J. K. Wolf, “Noiseless coding of correlated information sources,” *IEEE Trans. Inform. Theory*, vol. 19, no. 4, pp. 471–480, Apr. 1973. [1](#), [9](#), [10](#)
- [6] A. Wyner and J. Ziv, “The rate-distortion function for source coding with side information at the decoder,” *IEEE Trans. Inform. Theory*, vol. 22, no. 1, pp. 1–10, Jan. 1976. [1](#), [11](#)
- [7] T. Wiegand, G. Sullivan, G. Bjontegaard, and A. Luthra, “Overview of the h. 264/avc video coding standard,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 560–576, 2003. [2](#), [13](#)

- [8] C. Shannon, “A mathematical theory of communication,” *Bell Syst. Tech. J.*, vol. 27. 3
- [9] R. Zamir, S. Shamai, and U. Erez, “Nested linear/lattice codes for structured multiterminal binning,” *IEEE Trans. Inform. Theory*, vol. 48, no. 6, pp. 1250–1276, June 2002. 3, 23, 29, 47
- [10] Z. Liu, S. Cheng, A. D. Liveris, and Z. Xiong, “Slepian-Wolf coded nested quantization (SWC-NQ) for Wyner-Ziv coding: performance analysis and code design,” in *Proc. IEEE Data Compression Conf. (DCC)*, Mar. 2004, pp. 322–331. 5, 22, 29, 31, 32, 33, 43
- [11] D. Rebollo-Monedero, S. Rane, and B. Girod, “Wyner-Ziv quantization and transform coding of noisy sources at high rates,” in *Proc. Conference Record of the Thirty-Eighth Asilomar Conference on Signals, Systems and Computers*, 2004, pp. 2084–2088 Vol.2. 3, 23, 47
- [12] C. Brites, J. Ascenso, and F. Pereira, “Studying temporal correlation noise modeling for pixel based wyner-ziv video coding,” in *IEEE International Conference on Image Processing (ICIP) 2006*, Oct. 2006, pp. 273–276. 4, 65
- [13] C. Brites and F. Pereira, “Correlation noise modeling for efficient pixel and transform domain Wyner-Ziv video coding,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 9, pp. 1177–1190, Sept. 2008. 22, 48
- [14] X. Fan, O. C. Au, and N. M. Cheung, “Transform-domain adaptive correlation estimation (TRACE) for Wyner-Ziv video coding,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 11, pp. 1423–1436, nov. 2010. 4, 22, 65, 66, 70, 80
- [15] B. Girod, A. Aaron, S. Rane, and D. Rebollo-Monedero, “Distributed video coding,” *Proc. IEEE*, vol. 93, no. 1, pp. 71–83, June 2005. 4, 12, 51, 88, 90
- [16] Q. Xu, V. Stankovic, A. Liveris, and Z. X. Xiong, “Distributed joint source-channel coding of video,” in *Proc. IEEE International Conference on Image*

- Processing, (ICIP '05)*, Genova, Italy, Sept. 2005, pp. II-674–7 vol.2. [4](#), [23](#), [29](#), [35](#), [88](#)
- [17] L. A. Dalton, “Analysis of 1-D nested lattice quantization and Slepian-Wolf coding for Wyner-Ziv coding of i.i.d. sources,” Texas A&M University,” Project report for ELEN 663, May 2003. [5](#), [29](#), [31](#), [32](#), [41](#), [43](#), [48](#), [52](#), [58](#)
- [18] C. Shannon, “A mathematical theory of communication,” *Bell Syst. Tech. J.*, vol. 27, pp. 379–423, Oct. 1948. [10](#)
- [19] A. D. Wyner, “The rate-distortion function for source coding with side information at the decoder - II: General sources,” *Information and Control*, vol. 38, no. 1, pp. 60–80, 1978. [11](#)
- [20] M. Costa, “Writing on dirty paper,” *IEEE Trans. Inform. Theory*, vol. 29, no. 3, pp. 439–441, 1983. [11](#)
- [21] R. Barron, B. Chen, and G. W. Wornell, “The duality between information embedding and source coding with side information and some applications,” *IEEE Trans. Inform. Theory*, vol. 49, no. 5, pp. 1159–1180, 2003. [11](#)
- [22] R. Puri and K. Ramchandran, “Prism: A new robust video coding architecture based on distributed compression principles,” in *40th Allerton Conference on Communication, Control and Computing*, Allerton, IL, Oct. 2002. [12](#), [14](#)
- [23] C. Guillemot, F. Pereira, L. Torres, T. Ebrahimi, R. Leonardi, and J. Ostermann, “Distributed monoview and multiview video coding,” *IEEE Signal Processing Mag.*, vol. 34, no. 5, pp. 67–76, Sept. 2007. [12](#)
- [24] W. Liu, L. Dong, and W. Zeng, “Motion refinement based progressive side-information estimation for wyner-ziv video coding,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 12, pp. 1863–1875, dec. 2010. [14](#), [20](#), [21](#), [67](#), [68](#), [111](#)
- [25] Y. Zhang, H. Xiong, Z. He, S. Yu, and C. W. Chen, “An error resilient video coding scheme using embedded Wyner-Ziv description with decoder

- side non-stationary distortion modeling,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 21, no. 4, pp. 498–512, 2011. [15](#), [25](#)
- [26] M. W. Marcellin and T. R. Fischer, “Trellis coded quantization of memoryless and gauss-markov sources,” *IEEE Trans. on Comm.*, vol. 38, no. 1, pp. 82–93, 1990. [17](#)
- [27] S. S. Pradhan and K. Ramchandran, “Distributed source coding: Symmetric rates and applications to sensor networks,” in *Proc. IEEE Data Compression Conf. (DCC)*, 2000, pp. 363–372. [17](#)
- [28] —, “Group-theoretic construction and analysis of generalized coset codes for symmetric/asymmetric distributed source coding,” in *Proc. IEEE Data Compression Conf. (DCC)*, 2000.
- [29] —, “Geometric proof of rate-distortion function of Gaussian sources with side information at the decoder,” in *IEEE International Symposium on Information Theory*, 2000, pp. 351–351.
- [30] S. S. Pradhan, J. Kusuma, and K. Ramchandran, “Distributed compression in a dense microsensor network,” *IEEE Signal Processing Mag.*, vol. 19, no. 2, pp. 51–60, 2002.
- [31] X. Wang and M. T. Orchard, “Design of trellis codes for source coding with side information at the decoder,” in *Proc. IEEE Data Compression Conf. (DCC)*, 2001, pp. 361–370. [17](#)
- [32] J. Garcia-Frias, “Compression of correlated binary sources using turbo codes,” *IEEE Commun. Lett.*, vol. 5, no. 10, pp. 417–419, Oct. 2001. [18](#), [21](#), [29](#)
- [33] J. Garcia-Frias and Y. Zhao, “Data compression of unknown single and correlated binary sources using punctured turbo codes,” in *Allerton Conf. Communication, Control, and Computing*, vol. 39, no. 2, 2001, pp. 1282–1291.

- [34] J. Bajcsy and P. Mitran, “Coding for the Slepian-Wolf problem with turbo codes,” in *IEEE Global Telecommunications Conference(GLOBECOM’01)*, vol. 2, 2001, pp. 1400–1404.
- [35] A. Aaron and B. Girod, “Compression with side information using turbo codes,” in *Proc. IEEE Data Compression Conf. (DCC)*, UT, Snowbird, Apr. 2002, pp. 252–261. [29](#)
- [36] Y. Zhao and J. Garcia-Frias, “Joint estimation and compression of correlated nonbinary sources using punctured turbo codes,” *IEEE Trans. Commun.*, vol. 53, no. 3, pp. 385–390, 2005. [18](#)
- [37] —, “Data compression of correlated non-binary sources using punctured turbo codes,” in *Proc. IEEE Data Compression Conf. (DCC)*, 2002, pp. 242–251. [18](#)
- [38] P. Mitran and J. Bajcsy, “Coding for the wyner-ziv problem with turbo-like codes,” in *Proc. IEEE Int. Symp. Information Theory 2002*, 2002, p. 91. [18](#)
- [39] A. Liveris, Z. Xiong, and C. Georghiades, “Compression of binary sources with side information at the decoder using LDPC codes,” *IEEE Commun. Lett.*, vol. 6, no. 10, pp. 440–442, Oct. 2002. [18](#), [21](#), [29](#), [95](#)
- [40] —, “Compression of binary sources with side information using low-density parity-check codes,” in *IEEE Global Telecommunications Conference, 2002. GLOBECOM’02.*, vol. 2, 2002, pp. 1300–1304.
- [41] V. Stankovic, A. Liveris, Z. Xiong, and C. N. Georghiades, “Design of Slepian-Wolf codes by channel code partitioning,” in *Proc. IEEE Data Compression Conf. (DCC)*, 2004, pp. 302–311. [18](#)
- [42] D. Schonberg, K. Ramchandran, and S. Pradhan, “Distributed code constructions for the entire Slepian-Wolf rate region for arbitrarily correlated sources,” in *Proc. IEEE Data Compression Conf. (DCC)*, 2004, pp. 292–301. [18](#)

- [43] T. P. Coleman, A. H. Lee, M. Médard, and M. Effros, “On some new approaches to practical slepian-wolf compression inspired by channel coding,” in *Proc. IEEE Data Compression Conf. (DCC)*, 2004, pp. 282–291.
- [44] C. F. L. A. D. Liveris, K. Narayanan, Z. Xiong, and C. Georgiades, “Slepian-Wolf coding of multiple M-ary sources using ldpc codes,” in *Proc. IEEE Data Compression Conf. (DCC)*, 2004, p. 549.
- [45] N. Gehrig and P. L. Dragotti, “Symmetric and asymmetric slepian-wolf codes with systematic and nonsystematic linear codes,” *IEEE Commun. Lett.*, vol. 9, no. 1, pp. 61–63, 2005. [18](#)
- [46] D. Varodayan, A. Aaron, and B. Girod, “Rate-adaptive codes for distributed source coding,” *Signal Processing (EURASIP)*, no. 86, pp. 3123–3130, Nov. 2006. [18](#), [29](#), [55](#), [82](#), [98](#), [103](#)
- [47] M. Grangetto, E. Magli, and G. Olmo, “Distributed arithmetic coding,” *IEEE Commun. Lett.*, vol. 11, no. 11, pp. 883–885, 2007. [18](#)
- [48] —, “Distributed arithmetic coding for the slepian-wolf problem,” *IEEE Trans. Signal Processing*, vol. 57, no. 6, pp. 2245–2257, 2009.
- [49] —, “Symmetric distributed arithmetic coding of correlated sources,” in *IEEE 9th Workshop on Multimedia Signal Processing (MMSP)*, 2007, pp. 111–114.
- [50] M. Grangetto, E. Magli, R. Tron, and G. Olmo, “Rate-compatible distributed arithmetic coding,” *IEEE Commun. Lett.*, vol. 12, no. 8, pp. 575–577, 2008.
- [51] S. Malinowski, X. Artigas, C. Guillemot, and L. Torres, “Distributed coding using punctured quasi-arithmetic codes for memory and memoryless sources,” *IEEE Trans. Signal Processing*, vol. 57, no. 10, pp. 4154–4158, 2009.
- [52] M. Grangetto, E. Magli, and G. Olmo, “Decoder-driven adaptive distributed arithmetic coding,” in *IEEE International Conference on Image Processing (ICIP)*, 2008, pp. 1128–1131.

- [53] Y. Fang, “Distribution of distributed arithmetic codewords for equiprobable binary sources,” *IEEE Signal Processing Lett.*, vol. 16, no. 12, pp. 1079–1082, 2009. [18](#)
- [54] R. Puri and K. Ramchandran, “Prism: an uplink-friendly multimedia coding paradigm,” in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing, (ICASSP '03)*, Hong Kong, Apr. 2003, pp. IV–856–9 vol.4. [19](#)
- [55] A. Aaron, S. Rane, and B. Girod, “Wyner-Ziv video coding with hash-based motion compensation at the receiver,” in *Proc. IEEE International Conference on Image Processing, (ICIP '04)*, Singapore, Oct. 2004, pp. 3097–3100 Vol. 5. [19](#), [22](#), [90](#)
- [56] Z. Li and E. J. Delp, “Wyner-Ziv video side estimator: conventional motion search methods revisited,” in *IEEE International Conference on Image Processing (ICIP) 2005*, vol. 1, 2005, pp. I–825. [20](#)
- [57] A. Aaron, S. Rane, E. Setton, and B. Girod, “Transform-domain Wyner-Ziv codec for video,” in *Proc. SPIE Visual Communications and Image Processing*, vol. 5308, 2004, pp. 520–528. [20](#), [22](#)
- [58] X. Artigas, J. Ascenso, M. Dalai, S. Klomp, D. Kubasov, and M. Ouaret, “The discover codec: architecture, techniques and evaluation,” in *Picture Coding Symposium*, vol. 2007. Citeseer, 2007. [20](#)
- [59] D. Kubasov and C. Guillemot, “Mesh-based motion-compensated interpolation for side information extraction in distributed video coding,” in *IEEE International Conference on Image Processing*, 2006, pp. 261–264. [20](#)
- [60] G. Petrazzuoli, M. Cagnazzo, and B. Pesquet-Popescu, “High order motion interpolation for side information improvement in dvc,” in *IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)*, 2010, pp. 2342–2345. [20](#)
- [61] M. Akinola, L. Dooley, and P. Wong, “Wyner-ziv side information generation using a higher order piecewise trajectory temporal interpolation

- algorithm,” in *International Conference on Graphic and Image Processing*, 2010. [20](#)
- [62] R. Martins, C. Brites, J. Ascenco, and F. Pereira, “Refining side information for improved transform domain Wyner-Ziv video coding,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 19, no. 9, pp. 1327–1341, 2009. [20](#)
- [63] B. Macchiavello, D. Mukherjee, and R. L. de Queiroz, “Iterative side-information generation in a mixed resolution Wyner-Ziv framework,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 19, no. 10, pp. 1409–1423, 2009.
- [64] W. Liu, L. Dong, and W. Zeng, “Estimating side-information for Wyner-Ziv video coding using resolution-progressive decoding and extensive motion exploration,” in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing, 2009. (ICASSP 2009)*, Apr. 2009, pp. 721–724. [20](#)
- [65] —, “Wyner-Ziv video coding with multi-resolution motion refinement: theoretical analysis and practical significance,” in *Proc. SPIE Visual Communications and Image Processing (VCIP)*, 2008. [21](#)
- [66] X. Artigas and L. Torres, “Iterative generation of motion-compensated side information for distributed video coding,” in *IEEE International Conference on Image Processing, 2005. (ICIP) 2005.*, vol. 1, 2005, pp. I–833. [20](#)
- [67] Z. Li, “New methods for motion estimation with applications to low complexity video compression,” Ph.D. dissertation, Purdue University, 2005. [20](#)
- [68] J. Ascenco, C. Brites, and F. Pereira, “Motion compensated refinement for low complexity pixel based distributed video coding,” in *IEEE Conference on Advanced Video and Signal Based Surveillance, 2005. (AVSS 2005).*, 2005, pp. 593–598. [21](#)
- [69] B. Macchiavello, R. D. Queiroz, and D. Mukherjee, “Motion-based side-information generation for a scalable wyner-ziv video coder,” in *IEEE International Conference on Image Processing, 2007. (ICIP 2007).*, vol. 6, 2007, pp. VI–413. [20](#), [21](#)

- [70] D. Kubasov, J. Nayak, and C. Guillemot, “Optimal reconstruction in Wyner-Ziv video coding with multiple side information,” in *IEEE 9th Workshop on Multimedia Signal Processing (MMSP) 2007*, 2007, pp. 183–186. [22](#), [23](#), [46](#)
- [71] X. Guo, Y. Lu, F. Wu, D. Zhao, and W. Gao, “Wyner-ziv-based multiview video coding,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 6, pp. 713–724, 2008. [22](#), [67](#)
- [72] N. M. Cheung and A. Ortega, “Flexible video decoding: A distributed source coding approach,” in *IEEE 9th Workshop on Multimedia Signal Processing, 2007. (MMSP 2007).*, 2007, pp. 103–106. [22](#), [26](#)
- [73] N. M. Cheung, H. Wang, and A. Ortega, “Sampling-based correlation estimation for distributed source coding under rate and complexity constraints,” *IEEE Trans. Image Processing*, vol. 17, no. 11, pp. 2122–2137, 2008. [22](#), [24](#)
- [74] C. Brites, J. Ascenso, and F. Pereira, “Studying temporal correlation noise modeling for pixel based Wyner-Ziv video coding,” in *IEEE International Conference on Image Processing*, 2006, pp. 273–276. [22](#)
- [75] A. Aaron and B. Girod, “Wyner-Ziv video coding with low-encoder complexity,” in *Proc. Picture Coding Symposium*, 2004. [22](#)
- [76] C. Brites, J. Ascenso, and F. Pereira, “Modeling correlation noise statistics at decoder for pixel based Wyner-Ziv video coding,” in *Proc. Picture Coding Symposium*, 2006. [22](#)
- [77] M. Dalai, R. Leonardi, and F. Pereira, “Improving turbo codec integration in pixel-domain distributed video coding,” in *IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 2, 2006, pp. II–II. [22](#)
- [78] S. Lloyd, “Least squares quantization in pcm,” *IEEE Trans. Inform. Theory*, vol. 28, no. 2, pp. 129–137, 1982. [23](#)

- [79] M. Fleming, Q. Zhao, and M. Effros, “Network vector quantization,” *IEEE Trans. Inform. Theory*, vol. 50, no. 8, pp. 1584–1604, 2004.
- [80] D. Muresan and M. Effros, “Quantization as histogram segmentation: globally optimal scalar quantizer design in network systems,” in *Proc. IEEE Data Compression Conf. (DCC)*, 2002, pp. 302–311.
- [81] M. Effros and D. Muresan, “Codecell contiguity in optimal fixed-rate and entropy-constrained network scalar quantizers,” in *Proc. IEEE Data Compression Conf. (DCC)*, 2002, pp. 312–321.
- [82] J. Cardinal and G. V. Assche, “Joint entropy-constrained multiterminal quantization,” in *IEEE International Symposium on Information Theory*, 2002, p. 63. [23](#)
- [83] Y. Zhang and C. Zhu, “Quantizer design for correlation noise in distributed video coding,” in *IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*, 2011, pp. 1–6. [23](#)
- [84] S. D. Servetto, “Lattice quantization with side information: Codes, asymptotics, and applications in sensor networks,” *IEEE Trans. Inform. Theory*, vol. 53, no. 2, pp. 714–731, Feb. 2007. [23](#), [29](#)
- [85] J. Slowack, S. Mys, J. Skorupa, N. Deligiannis, P. Lambert, A. Munteanu, and R. V. de Walle, “Rate-distortion driven decoder-side bitplane mode decision for distributed video coding,” *Signal Processing: Image Communication*, vol. 25, no. 9, pp. 660–673, 2010. [23](#)
- [86] Y. Zhang, H. Xiong, Z. He, and S. Yu, “Reconstruction for distributed video coding: a markov random field approach with context-adaptive smoothness prior,” in *Proceedings of SPIE*, vol. 7744, 2010, p. 774417. [23](#)
- [87] M. Morbee, A. Roca, J. Prades-Nebot, A. Pižurica, and W. Philips, “Reduced decoder complexity and latency in pixel-domain Wyner-Ziv video coders,” *Signal, Image and Video Processing*, vol. 2, no. 2, pp. 129–140, 2008. [24](#)

- [88] D. Kubasov, K. Lajnef, and C. Guillemot, “A hybrid encoder/decoder rate control for wyner-ziv video coding with a feedback channel,” in *IEEE 9th Workshop on Multimedia Signal Processing, 2007. (MMSP 2007).*, 2007, pp. 251–254.
- [89] J. Areia, J. Ascenso, C. Brites, and F. Pereira, “Low complexity hybrid rate control for lower complexity wyner-ziv video decoding,” in *Proceedings of European Conference on Signal Processing (EUSIPCO08)*, 2008. [24](#)
- [90] M. Morbée, J. Prades-Nebot, A. Pizurica, and W. Philips, “Rate allocation algorithm for pixel-domain distributed video coding without feedback channel,” in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2007*, vol. 1. IEEE, 2007, pp. I-521. [24](#)
- [91] W. Weerakkody, W. Fernando, and A. Adikari, “Unidirectional distributed video coding for low cost video encoding,” *IEEE Trans. Consumer Electron.*, vol. 53, no. 2, pp. 788–795, 2007.
- [92] C. Brites and F. Pereira, “Encoder rate control for transform domain wyner-ziv video coding,” in *IEEE International Conference on Image Processing (ICIP) 2007*, vol. 2, 2007, pp. II-5.
- [93] J. Martínez, G. Fernández-Escribano, H. Kalva, W. Weerakkody, W. Fernando, and A. Garrido, “Feedback free DVC architecture using machine learning,” in *15th IEEE International Conference on Image Processing (ICIP) 2008*, 2008, pp. 1140–1143.
- [94] T. Sheng, G. Hua, H. Guo, J. Zhou, and C. W. Chen, “Rate allocation for transform domain wyner-ziv video coding without feedback,” in *Proceeding of the 16th ACM international conference on Multimedia*, 2008, pp. 701–704. [24](#)
- [95] A. Aaron, S. Rane, D. Rebollo-Monedero, and B. Girod, “Systematic lossy forward error protection for video waveforms,” in *IEEE International Conference on Image Processing*, vol. 1, 2003, pp. I-609. [25](#)

- [96] S. Rane, P. Baccichet, and B. Girod, “Systematic lossy error protection of video signals,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 10, pp. 1347–1360, 2008. [25](#)
- [97] Y. Zhang, C. Zhu, and K. H. Yap, “A joint source-channel video coding scheme based on distributed source coding,” *IEEE Trans. Multimedia*, vol. 10, no. 8, pp. 1648–1656, Dec. 2008. [25](#), [55](#)
- [98] Y. Wang, A. R. Reibman, and S. Lin, “Multiple description coding for video delivery,” *Proc. IEEE*, vol. 93, no. 1, pp. 57–70, 2005. [25](#)
- [99] A. Jagmohan, A. Sehgal, and N. Ahuja, “Two-channel predictive multiple description coding,” in *IEEE International Conference on Image Processing*, vol. 2, 2005, pp. II–670. [25](#)
- [100] O. Crave, B. Pesquet-Popescu, and C. Guillemot, “Robust video coding based on multiple description scalar quantization with side information,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 6, pp. 769–779, 2010. [25](#)
- [101] Y. Fan, J. Wang, and J. Sun, “Distributed multiple description video coding on packet loss channels,” *IEEE Trans. Image Processing*, no. 99, 2011. [25](#)
- [102] Y. Zhang, C. Zhu, and Y. Xu, “Multiple description wyner-ziv video coding based on decoder multi-resolution motion refinement,” in *IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*. IEEE, 2012, pp. 1–6. [25](#)
- [103] N. M. Cheung and A. Ortega, “Compression algorithms for flexible video decoding,” in *Visual Communications and Image Processing*, vol. 6822, no. 1, 2008. [26](#)
- [104] J. Park and B. Jeon, “Parallel transform domain wyner-ziv video encoding scheme for standard definition video,” in *IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*. IEEE, 2012, pp. 1–4. [27](#)

- [105] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley-Interscience, 1991. [28](#), [95](#)
- [106] A. Wyner, “Recent results in the Shannon theory,” *IEEE Trans. Inform. Theory*, vol. 20, no. 1, pp. 2–10, Jan. 1974. [29](#)
- [107] G. D. Forney, “Coset codes part I: Introduction and geometric classification,” *IEEE Trans. Inform. Theory*, vol. 34, no. 5, pp. 1123–1151, Sept. 1988. [29](#), [92](#)
- [108] —, “Coset codes part II: Binary lattices and related codes,” *IEEE Trans. Inform. Theory*, vol. 34, no. 5, pp. 1152–1187, Sept. 1988. [92](#)
- [109] J. H. Conway and N. J. A. Sloane, *Sphere Packings, Lattices and Groups*, 3rd ed. New York: Springer-Verlag, 1998. [29](#)
- [110] J. Ascenso, C. Brites, and F. Pereira, “Motion compensated refinement for low complexity pixel based distributed video coding,” in *IEEE International Conference on Advanced Video and Signal Based Surveillance*, Como, Italy, Sept. 2005. [43](#)
- [111] J. Ascenso, C. Brites, and F. Pereira, “Content adaptive Wyner-Ziv video coding driven by motion activity,” in *Proc. International Conference on Image Processing*, Atlanta, GA, Oct. 2006. [43](#), [81](#)
- [112] R. Westerlaken, R. K. Gunnewiek, and R. Lagendijk, “Turbo-code based Wyner-Ziv video compression,” in *Twenty-sixth Symposium on Information Theory*, Benelux, May 2005, pp. 113–120. [48](#)
- [113] A. Avudainayagam, J. Shea, and D. Wu, “Hyper-trellis decoding of pixel-domain Wyner-Ziv video coding,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 5, pp. 557–568, May 2008. [62](#)
- [114] V. Toto-Zarasoá, A. Roumy, and C. Guillemot, “Source modeling for distributed video coding,” *IEEE Trans. Circuits Syst. Video Technol.*, no. 99, 2011. [66](#)

- [115] Z. Xue, K. K. Loo, J. Cosmas, M. Tun, L. Feng, and P. Y. Yip, "Error-resilient scheme for wavelet video codec using automatic roi detection and wyner-ziv coding over packet erasure channel," *IEEE Trans. Broadcast.*, vol. 56, no. 4, pp. 481–493, 2010. [67](#)
- [116] T. A. Summers and S. G. Wilson, "Snr mismatch and online estimation in turbo decoding," *IEEE Trans. Commun.*, vol. 46, no. 4, pp. 421–423, 1998. [70](#)
- [117] S. Cheng and Z. Xiong, "Successive refinement for the wyner-ziv problem and layered code design," *IEEE Trans. Signal Processing*, vol. 53, no. 8, pp. 3269–3281, 2005. [70](#), [71](#), [78](#)
- [118] D. S. Taubman and M. W. Marcellin, "Jpeg2000: Image compression fundamentals, standards and practice," 2002. [71](#)
- [119] J. M. Shapiro, "Embedded image coding using zerotrees of wavelet coefficients," *IEEE Trans. Signal Processing*, vol. 41, no. 12, pp. 3445–3462, 1993. [72](#)
- [120] J. R. Ohm, "Wavelet video coding," in *International workshop on systems, signals and image processing, ambient multimedia*, 2004, pp. 1–3. [73](#)
- [121] Z. Li, L. Liu, and E. J. Delp, "Rate distortion analysis of motion side estimation in Wyner-Ziv video coding," *IEEE Trans. Image Processing*, vol. 16, no. 1, pp. 98–113, Jan. 2007. [81](#)
- [122] Y. Wang, J. Ostermann, and Y.-Q. Zhang, *Video Processing and Communications*. Prentice Hall, 2002. [88](#)
- [123] W.-T. Tan and A. Zakhor, "Video multicast using layered FEC and scalable compression," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, no. 3, pp. 373–386, Mar. 2001. [88](#)
- [124] B. Kurceren and J. W. Modestino, "A joint source-channel coding approach to network transport on digital video," in *Proc. IEEE Nineteenth Annual Joint Conference of the Computer and Communications Societies, INFO-COM 2000*, vol. 2, Mar. 2000, pp. 717–726. [88](#)

- [125] A. D. Liveris, Z. X. Xiong, and C. N. Georgiades, “Joint source-channel coding of binary sources with side information at the decoder using IRA codes,” in *IEEE Workshop on Multimedia Signal Processing*, Dec. 2002. [89](#), [92](#)
- [126] Q. Xu, V. Stankovic, and Z. Xiong, “Distributed joint source-channel coding of video using raptor codes,” *IEEE Journal on Selected Areas in Communications*, vol. 25, no. 4, pp. 851–861, 2007. [89](#)
- [127] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, “Overview of the H.264/AVC video coding standard,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 560–576, July 2003. [90](#)
- [128] A. Aaron, D. Varodayan, and B. Girod, “Wyner-Ziv residual coding of video,” in *Proc. IEEE International Picture Coding Symposium, (IPCS '06)*, Beijing, P. R. China, Apr. 2006. [90](#)