LAYERED WYNER-ZIV VIDEO CODING FOR NOISY CHANNELS

A Thesis

by

QIAN XU

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

August 2004

Major Subject: Electrical Engineering

LAYERED WYNER-ZIV VIDEO CODING

FOR NOISY CHANNELS

A Thesis

by

QIAN XU

Submitted to Texas A&M University
in partial fulfillment of the requirements
for the degree of

MASTER OF SCIENCE

Approved as to style and content by:

Zixiang Xiong (Chair of Committee)	
Costas N. Georghiades (Member)	Andrew K. Chan (Member)
Dmitri Loguinov (Member)	Chanan Singh (Head of Department)

August 2004

Major Subject: Electrical Engineering

ABSTRACT

Layered Wyner-Ziv Video Coding for Noisy Channels. (August 2004)

Qian Xu, B.S., University of Science & Technology of China

Chair of Advisory Committee: Dr. Zixiang Xiong

The growing popularity of video sensor networks and video celluar phones has generated the need for low-complexity and power-efficient multimedia systems that can handle multiple video input and output streams. While standard video coding techniques fail to satisfy these requirements, distributed source coding is a promising technique for "uplink" applications. Wyner-Ziv coding refers to lossy source coding with side information at the decoder. Based on recent theoretical result on successive Wyner-Ziv coding, we propose in this thesis a practical *layered* Wyner-Ziv video codec using the DCT, nested scalar quantizer, and irregular LDPC code based Slepian-Wolf coding (or lossless source coding with side information) for noiseless channel. The DCT is applied as an approximation to the conditional KLT, which makes the components of the transformed block conditionally independent given the side information. NSQ is a binning scheme that facilitates layered bit-plane coding of the bin indices while reducing the bit rate. LDPC code based Slepian-Wolf coding exploits the correlation between the quantized version of the source and the side information to achieve further compression. Different from previous works, an attractive feature of our proposed system is that video encoding is done only once but decoding allowed at many lower bit rates without quality loss. For Wyner-Ziv coding over discrete noisy channels, we present a Wyner-Ziv video codec using IRA codes for Slepian-Wolf coding based on the idea of two equivalent channels. For video streaming applications where the channel is packet based, we apply unequal error protection scheme to the embedded Wyner-Ziv coded video stream to find the optimal source-channel coding

trade-off for a target transmission rate over packet erasure channel.

To my wonderful parents and my dear sister

ACKNOWLEDGMENTS

I would first like to thank my advisor, Dr. Zixiang Xiong, for introducing me to distributed video coding. He has been an excellent guide in my transition from undergraduate school to graduate school. Without his support, advice and constant encouragement, this work would not have been possible. I also thank my other committee members, Dr. Costas N. Georghiades, Dr. Andrew K. Chan, Dr. Dmitri Loguinov, for their valuable comments and for committing their time to help me with this work. Thanks also to the remaining professors of the wireless communications group: Dr. Scott Miller, Dr. Krishna Narayanan, Dr. Erchin Serpedin, and Dr. Deepa Kundur, for their instructions and teaching, inside and output the classroom.

I want to express my sincere gratitude to my colleagues in the Multimedia Laboratory, for sharing their insightful knowledge with me. I am especially grateful to Jianping Hua, Samuel Cheng, Vladimir Stanković, Yang Yang and Tim Lan for their many helpful suggestions and assistance.

Finally, I would like to thank my family. I would not be where I am today without the unconditional love, support and encouragement of my parents and sister.

TABLE OF CONTENTS

CHAPTER Page

LIST OF TABLES

LIST OF FIGURES

FIGURE Page

CHAPTER I

INTRODUCTION

Today's standard techniques for video compression are designed for "downlink" broadcast applications with one heavy encoder and multiple light decoders. Video coding standards like MPEG [8] and H.26X [9] use motion-compensated predictive DCT to achieve high compression efficiency. The encoder is the computational workhorse of the video codec while the decoder is a relatively lightweight device operating in a "slave" mode. Therefore, they are suitable for video communications (e.g. broadcast) where encoding is done only once without any power constraint and decoding performed many times.

The growing popularity of video sensor networks, video cellular phones and webcams has generated the need for low-complexity and power-efficient multimedia systems that can handle multiple video input and output streams. For example, when a natural scene is captured by spatially separated cameras and transmitted over noisy channels to a central base station for decoding, a typical new scenario of "uplink" multimedia applications arises, which has very different requirements from the traditional "downlink" scenarios. For such applications, we need a video coding system with multiple low-complexity encoders and one (or more) high-complexity decoders. In addition, the system must be robust to channel errors so that the decoder at the base station can recover the scene with high fidelity using all received bitstreams.

While standard video coding techniques (e.g., MPEG [8] and H.26X [9]) provide high compression efficiency, they fail to satisfy the requirements of the above "uplink" multimedia application. This is because the heavy computation load of DCT and

---

The journal model is *IEEE Transactions on Automatic Control.*

motion estimation is put at the encoder while the decoder is a relatively lightweight device. Typically, the complexity of a standard encoder is 5 to 10 times higher than that of the decoder. Moreover, when there are channel errors or packet losses, a decoded frame at the decoder will be different from that used at the encoder, causing the problem of error drifting that will have adverse effect on subsequent frames with severe visual degradation.

Distributed source coding (DSC) is a promising technique for "uplink" applications. DSC refers to the compression of the two or more separated sources that do not communicate with each other. Video coding based on DSC principles enables a new architecture with many encoders and one decoder that effectively swaps the encoder-decoder complexity of standard codecs. The encoders can just "blindly" compress the video inputs independently, while leaving the decoder to exploit the correlation among them. Therefore, by shifting the coding complexity from the encoder to the decoder, distributed video coding achieves an asymmetric structure that is the exact opposite of standard video coding.

With this new DSC paradigm, the problem is then how to achieve the same coding efficiency as traditional video coding. The problem of separate encoding and joint decoding of two correlated sources was first considered by Slepian and Wolf [10], who proved that there is no loss of coding efficiency with separate encoding when compared to joint encoding as long as joint decoding is performed. For the more general case of lossy coding with side information at the decoder, Wyner and Ziv [11] showed that it generally suffers rate loss when compared to lossy coding of the source with the side information available at both the encoder and the decoder. However, a special case of the Wyner-Ziv problem is when the source X and side information Y are zero mean and stationary Gaussian memoryless sources and the distortion metric is MSE. The minimum bit rate needed to encode X for a given distortion when Y

is only available at the decoder is equal to the rate when Y is known at both sides. In other words, there is no rate loss for the quadratic Gaussian case in Wyner-Ziv coding (WZC)!

To approach the Wyner-Ziv rate-distortion function established in [11], several practical coding schemes for ideal sources have been proposed [12][13][14][15][16][17][18]. Applying the Wyner-Ziv coding principle to video sources has only begun recently. Aaron *et al.* [1] proposed a distributed video compression scheme and addressed its error resilience property. However, the proposed systems incur a substantial rate-distortion (R-D) penalty compared to standard MPEG-4 coding. Sehgal *et al.* [2] discussed how coset-based Wyner-Ziv video coding can alleviate the problem of prediction mismatch. Puri and Ramchandran [3] outlined a PRISM framework that swaps the encoder/decoder complexity in standard codecs. Only performance at high bit rates was provided in [3].

In this thesis, we present a novel layered video coding scheme based on successive refinement for the Wyner-Ziv problem [4] and addressed its error robustness. Treating a standard coded video as the base layer (or side information), a layered Wyner-Ziv bitstream of the original video sequence is generated to enhance the base layer such that it is still decodable with commensurate qualities at rates corresponding to layer boundaries. Thus our proposed layered WZC scheme is very much like MPEG-4/H.26L FGS (Fine Granularity Scalable) coding [19][20] in "spirit" in terms of having an embedded enhancement layer with good R-D performance. However, the key difference is that the enhancement layer is generated "blindly" without knowing the base layer in WZC. This avoids the problems (e.g., error drifting/propagation) associated with encoder-decoder mismatch in standard DPCM-based coders. Using the H.26L coded version as the base layer, the proposed layered Wyner-Ziv video coding system over noiseless channel has roughly the same R-D performance as that

of H.26L-FGS [20] coding, with about 0.3dB Peak Signal-to-Noise Ratio (PSNR) loss at high rate.

Compared to the scheme in [1] that suffers a huge performance loss in the "embedded" mode, our layered Wyner-Ziv coding scheme has the attractive feature that encoding is done only once but decoding allowed at many lower bit rates with commensurate qualities and no quality loss. This is because our work is underpinned by recent theoretical results [4, 5] that extend the successive refinability of Gaussian sources from classic source coding [21] to WZC and because our design is based on scalar quantization and bit-plane coding. While the code design in [5] assumes ideal Gaussian or binary sources, results here are the first reported on practical layered WZC of video that *do not* suffer performance loss due to layering.

For discrete noisy channels, we first propose a *joint source-channel coding* (JSCC) framework for Wyner-Ziv video coding over a binary symmetric channel (BSC). A Wyner-Ziv coder can be thought of as a quantizer followed by Slepian-Wolf coding (SWC), which can be viewed as a channel coding problem. The channel coding component used for error protection can be combined with the SWC component in a joint design. Irregular Repeat Accumulate (IRA) codes [22][23], which are a special class of low-density parity-check (LDPC) codes [24], are employed for this purpose. The standard coded video (or base layer) can be viewed as the systematic part of the IRA code which is transmitted through the systematic channel and the Wyner-Ziv coded bitstream as the parity part transmitted through the real channel. Advanced design techniques (e.g. density evolution using Gaussian approximation [22]) can be employed to optimize the IRA code according to different conditions on both channels. By transmitting the video source over two channels, the error correcting ability of the joint source-channel code automatically protects the system from packet losses and errors. Compared with the traditional video transmission systems which

use feedback and retransmission of lost packets based on forward error correction (FEC), our system enables one integrated design of both the systematic part and the parity part of the IRA codes for different channel conditions and provides better R-D performance. In practical simulation, this practical Wyner-Ziv video coding system using IRA codes is only about 0.08 b/s away from the theoretical limit over the BSC with $p = 0.01$.

For video streaming applications where the transmission channel is packet based, a Wyner-Ziv video coding scheme using unequal error protection (UEP) [61] of embedded data for packet erasure channel is presented. UEP generates optimal packing scheme that minimizes the expected distortion at a target transmission rate. Consider the output bit planes after WZC as a embedded bitstream with decreasing importance from the most significant bit (MSB) to the least significant bit (LSB), UEP protects the encoded syndromes from packet losses in unreliable channels at a given transmission rate. With a combined design of WZC and UEP, we obtain an efficient video coding and transmission system over packet erasure channels. Simulations were carried out for some video sequences over packet erasure channel with packet mean loss rate 0.2. To achieve the same PSNR performance as in noiseless channel case, an extra 0.12 b/s is required to provide erasure protection for "Foreman" sequence and 0.16 b/s is required for "Mother_daughter" sequence.

The rest of the thesis is organized as follows: In Chapter II, we will introduce the theory of DSC, which covers SWC for distributed lossless compression and WZC for lossy source coding with side information at the decoder. Our practical layered Wyner-Ziv video coding scheme is presented in Chapter III using LDPC code based bit plane coding for SWC. In chapter IV, we consider the problem of video coding and transmission over noisy discrete channels. A practical Wyner-Ziv video coding framework based on JSCC with side information at the decoder using IRA codes is

discussed. For real-time applications where short delay and low memory are desired, UEP scheme is applied to the Wyner-Ziv coded bitstreams to provide protection from packet losses over packet erasure channels. The optimization of the UEP scheme and its performance are addressed in Chapter V. Finally conclusions are drawn in Chapter VI.

CHAPTER II

SOURCE CODING WITH SIDE INFORMATION (SCSI)

Although the theoretical foundation of DSC was laid by Slepian and Wolf [10] and Wyner and Ziv [11] in the early 1970s, research on practical code designs have started only recently. In the following, we briefly review theoretical results in SWC and WZC before addressing code designs.

A. Slepian-Wolf Coding: Theory

Consider $\{(X_i, Y_i)\}_{i=1}^{\infty}$ as a sequence of independent drawings of a pair of discrete random variables X, Y from a given distribution $p_{XY}(x, y)$ (see Fig. 1 (a)). Then for lossless compression of X and Y, a sum rate $R = R_X + R_Y = H(X, Y)$ will be sufficient if they are encoded jointly [25]. To perform joint encoding, both X and Y are assumed to be available at the encoder. A simple way to achieve this sum rate is to first encode X using $H(X)$ bits, then Y using $H(Y|X)$ bits based on the perfect knowledge of X. What if we have to meet the requirement of DSC that the two sources must be separately encoded (Fig. 1 (b))?

Such a problem of separate encoding and joint decoding of two correlated sources was first considered by Slepian and Wolf [10] and the answer is given in the achievable rate region defined as

$$R_X \geq H(X|Y), \qquad R_Y \geq H(Y|X), \qquad R_X + R_Y \geq H(X, Y) \tag{2.1}$$

which is shown in Fig. 2. This result is quite surprising in that the joint entropy $H(X, Y)$ is still achievable as long as the individual rate for each source is at least its conditional entropy given the other source. Therefore, there is no loss of coding efficiency with separate encoding when compared to joint encoding as long as joint

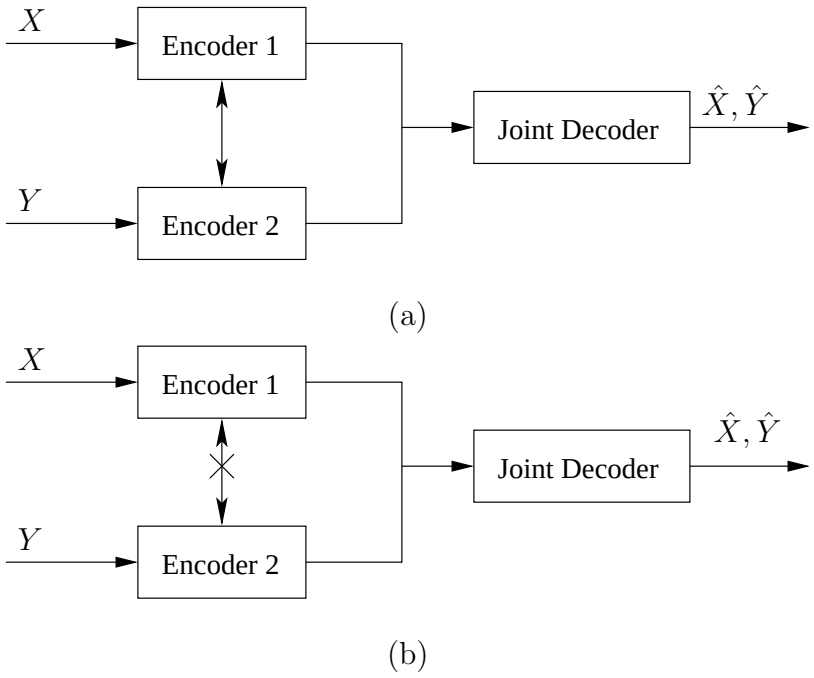


Fig. 1. Illustration of joint and distributed source coding. (a)Joint encoding of X and Y. $H(X,Y)$ is sufficient since the two encoders can collaborate. (b) Separate encoding of X and Y. Although the two encoders can not communicate with each other, Slepian-Wolf theorem states that $H(X,Y)$ is still sufficient for lossless recovery.

decoding is performed.

Lossless source coding with side information at the decoder, as shown in Fig. 3, is a special case of the SWC problem. Assume we give enough rate to Y for lossless recovery, Y is encoded using $H(Y)$ bits so that it can be perfectly decoded at the decoder. So the SWC problem boils down to compressing X to the rate limit $H(X|Y)$, i.e., achieving the corner point A in Fig. 2. Such a coding scheme is referred to as *asymmetric coding*. *Symmetric coding* (Fig. 1 (b)), on the other hand, aims to approach any point between A and B in the Slepian-Wolf rate region.

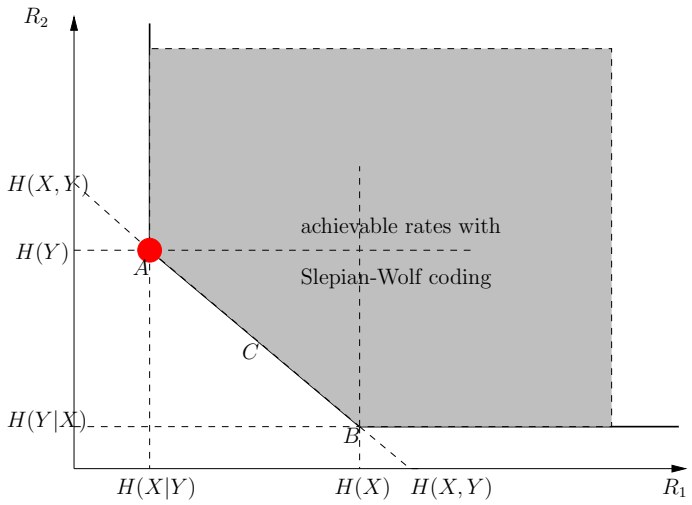


Fig. 2. Achievable rate region of Slepian-Wolf coding of two sources.

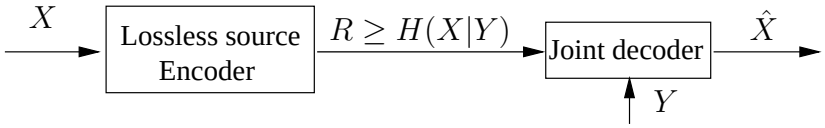


Fig. 3. Lossless source coding with side information at the decoder as a special case of Slepian-Wolf coding.

B. Slepian-Wolf Coding: Code Design for Ideal Sources

The proof of the Slepian-Wolf theorem is based on random binning, which is asymptotic and non-constructive, hence not applicable in practical code design. In practice, we may try first to design codes to approach the corner point A with $R_1 = H(X|Y)$ and $R_2 = H(Y)$ in Fig. 2. If this can be done, then the other corner point B can also be achieved by swapping the roles of X and Y. Then all the points between A and B can also be achieved through time sharing. In this thesis we will constrained ourselves to asymmetric code designs.

Wyner first suggested the use of parity-check codes for SWC in 1974 [26]. The

basic idea is to partition the codeword space into cosets using "good" parity-check code and then only transmit the index of the coset that the source code belongs to. The channel code should be "good" in the sense that the distance between codewords in the same coset should be as far as possible to facilitate decoding. Specifically, consider a linear (n, k) block parity-check code over $GF(2)$, which partitions the n dimensional vector space C^n into 2^k subspace C_j^n, $j = 1, \ldots, 2^{n-k}$, each of which contains 2^k codewords. The codewords c_{jm}, $m = 1, \ldots, 2^k$ among the same coset C_j^n share the property that $c_{jm} \times H^T = s_j$, $\forall m$, where H is the $(n-k) \times n$ parity check matrix of the code and s_j is a vector of length $n-k$ which denotes the *syndrome* of this coset. Different cosets correspond to different syndromes, and the coset with syndrome s_j equals to 0 is in fact the original channel code specified by H. In addition, the Hamming distance property of the channel code is preserved in each coset. In compressing, a sequence of n input bits is mapped to the $n-k$ syndrome bits that index the coset it belongs to, achieving a compression ratio of $n : (n-k)$. Using the coded coset index, the decoder finds in the coset the codeword closest to the side information as the best estimate of the input sequence. This approach, known as "Wyner's scheme" [26], means that channel codes can be used to perform compression in the Slepian-Wolf setup.

Wyner's above scheme was implemented in [16] based on traditional channel code like block and trellis code. If the correlation between X and Y can be modeled as a correlation channel, a good code for this channel will provide a good Slepian-Wolf code according to Wyner's syndrom-based scheme. Hence state-of-the-art capacity-achieving channel codes such as turbo [27] and LDPC [24] codes can be used to aproach the Slepian-Wolf limit. Practical designs based on turbo codes were reported in [28][29][30] by sending the parity bits of turbo codes instead of syndrom bits as advocated in Wyner's scheme. The first work that follow Wyner's scheme in devising

SWC schemes is presented in [31] using turbo codes and in [32] using LDPC codes. The reported performance in [31][32][33][34] are better than those in [28][29][30] and very close to the Slepian-Wolf limit.

C. Wyner-Ziv Coding: Theory

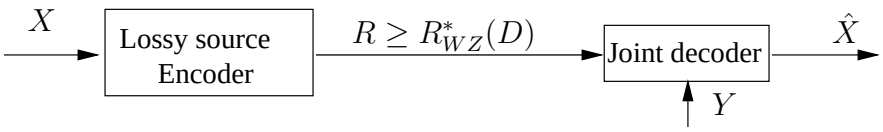


Fig. 4. Lossy source coding with side information at the decoder, i.e., Wyner-Ziv coding.

In the previous two sections, we addressed the problem of lossless source coding of discrete sources with side information at the decoder. In practical applications (e.g., distributed video coding and sensor networks), we deal with continuous sources and perform lossy coding rather than lossless coding. Hence we will extend SWC to lossy source coding with side information at the decoder. WZC, as depicted in Fig. 4, generalizes the setup of SWC in that coding of X is with respect to a fidelity criterion rather than lossless. So the question to ask is how many bits are needed to encode the source X under the constraint that the average distortion between X and decoded version $\hat{X}$ satisfies $E\{d(X, \hat{X})\} \leq D$, assuming that the side information Y is available only at the decoder. Denote $R^*_{WZ}(D)$ as the achievable lower bound for the bit-rate for an expected distortion D for WZC, and $R_{X|Y}(D)$ as the rate required if the side information is available also at the encoder.

In general there is a rate loss associated with WZC compared to the case when Y is also available at the encoder, as Wyner and Ziv proved that [11] $R^*_{WZ}(D) \geq R_{X|Y}(D)$. For example, when X and Y are binary symmetric sources and the corre-

lation between them can be modeled as a BSC with crossover probability p. Wyner and Ziv showed that [11], with Hamming distance measure, the rate-distortion function is:

$$R^*_{WZ}(D) = l.c.e\{H((1-p)D + (1-D)p) - H(D), (p, 0)\}, \quad x \leq D \leq p, \tag{2.2}$$

where $l.c.e.$ denotes the lower convex envelope. Note that when the encoder also has access to Y, the rate-distortion function becomes

$$R_{X|Y}(D) = \begin{cases} H(p) - H(D); & 0 \leq D \leq \min\{p, 1-p\}, \\ 0; & D > \min\{p, 1-p\}. \end{cases} \tag{2.3}$$

Therefore, $R^*_{WZ}(D) \geq R_{X|Y}(D)$, for $0 < D < p \leq 0.5$. Zamir *et. al.* showed that [35] the rate loss for binary sources with Hamming distance is less than 0.33 bit and for continuous sources with MSE measure is less than 0.5 b/s.

However, an exception occurs when X and Y are zero mean and jointly Gaussian and the distortion measure is MSE. Let the covariance matrix between X and Y be $\Lambda = \begin{bmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{bmatrix}$ with $|\rho| < 1$, then according to [11],

$$R^*_{WZ}(D) = R_{X|Y}(D) = \frac{1}{2}\log^+[\frac{\sigma_X^2(1-\rho^2)}{D}], \tag{2.4}$$

where $\log^+ x = \max\{\log x, 0\}$. So there is no rate loss in this quadratic Gaussian case. Recently Pradhan *et al.* [36] extended this no rate loss result to the more general case with $X = Y + Z$, where X and Y are independent and only Z is Gaussian (Y can follow arbitrary distribution). In this work, we limit ourselves to the quadratic Gaussian case because 1) There is no rate loss for WZC in the quadratic Gaussian case; 2) It is of special interest in practice because many image and video sources can be modeled as jointly Gaussian after mean subtraction.

D. Wyner-Ziv Coding: Code Design for Ideal Sources

As with SWC, efforts towards practical WZC have been undertaken only recently. In general, a Wyner-Ziv coder can be thought of as a quantizer (source code) followed by a Slepian-Wolf coder, as illustrated in Fig. 5.

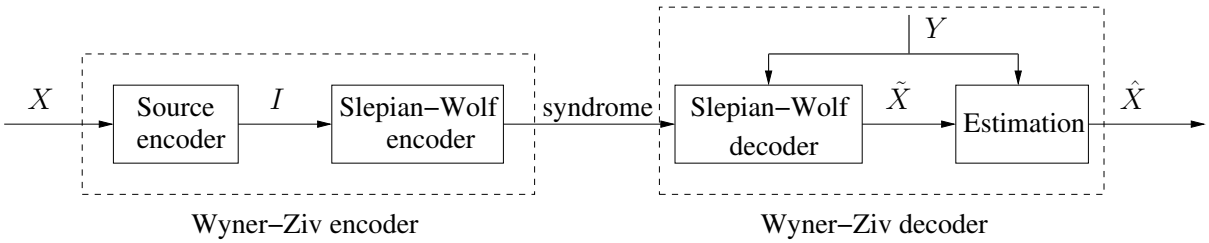


Fig. 5. Block diagram of a Wyner-Ziv coder.

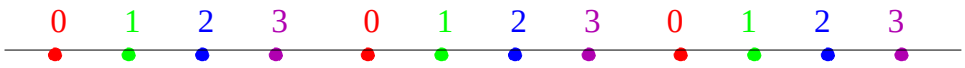


Fig. 6. 1-D nested lattice.

For example, Fig. 6 is a 1-D nested quantizer, which is a coarse coset code nested in a fine coset code, i.e., the coarse code is a subcode of the fine code. The fine code plays the role of source coding while each coarse coset code does channel coding. This coset coding scheme amounts to binning, which refers to partitioning the space of all possible outcomes of a random source into disjoint subsets or bins. To encode, X is first quantized by the fine source code, resulting in quantization errors, then only the index of the bin (coset) that the quantized X belongs to is coded to save rate. Using this coded index, the decoder finds in the bin (coset code) the codeword closest to the side information Y as the best estimate of X.

Usually, there is still correlation remaining in the quantized version of X and

the side information Y, and SWC should be employed to exploit this correlation to reduce rate. So WZC, in a nutshell, is a joint source-channel coding problem. There are quantization errors due to source coding and binning loss due to channel coding. To approach the Wyner-Ziv limit, one needs to employ both source codes (e.g., TCQ [37]) with high granular gain and channel code (e.g., turbo [27] and LDPC codes [24]) that can approach the Slepian-Wolf limit.

The simplest 1-D nested quantizer with $N = 4$ bins is shown in Fig. 6, where the fine source is a uniform scalar quantizer with stepsize q and the coarse channel code is a 1-D lattice code with minimum distance $d_{min} = Nq$. Two types of distortion are introduced: the quantizer incurs source coding error D_{sc}, which is $q^2/12$ at high rate, and the coarse channel code leads to channel coding error D_{cc}, which is inversely proportional to $d_{min} = Nq$. Hence for a fixed N, it is desirable to search for the optimal q that minimizes the total distortion $D = D_{sc} + D_{cc}$.

Lattice codes [38] and trellis-based codes [37] have been used in finding good nesting code in higher dimensions. Following the nested quantization scheme proposed by Zamir *et al.* in [12][39], Servetto [13] proposed explicit nested lattice constructions. Using trellis-based nested codes as a way of realizing high-dimensional nested lattice codes, Pradhan and Ramchandran [16] proposed DISCUS for WZC. Wang and Orchard [14] employed TCQ and trellis code for coding with side information. Chou *et al.* [15] used TCQ and turbo codes. A different approach to the practical Wyner-Ziv code design problem is Slepian-Wolf coded quantization (SWCQ) [17][40], which follows the reasoning that the classic entropy coder should be replaced by Wyner's syndrome-based Slepian-Wolf coder. The results in [17][40][18] show that the performance gap of high-rate SWCQ to the Wyner-Ziv distortion-rate function $D^*_{WZ}(R)$ is exactly the same as that of high-rate classic source coding to the distortion-rate function $D(R)$. In the practical design example [18] of 2-D TCVQ, irregular LDPC

codes based SWC, and optimal estimation at the decoder, the performance gap to the Wyner-Ziv limit is only 0.66 dB at 1.0 b/s and 0.47 dB at 3.3 b/s.

E. Successive Wyner-Ziv Coding: Theory

The problem of successive refinement of information was originally formulated by Equitz and Cover [21]. A source X is to be encoded and transmitted through a rate-limited channel. With rate R_1, the decoder produces $\hat{X}_1$, which is an approximation of X as distortion level D_1. At a later stage, the encoder sends a secondary string at rate ΔR to the decoder. With both bitstreams at hand, the decoder will produce $\hat{X}_2$, a more accurate reconstruction of X at distortion level D_2. If successive coding in two or more stages can be made R-D optimal simultaneously at all stages, the source is called *successively refinable*. For the two-stage case, the two rates should simultaneously lie on the R-D curve, i.e.,

$$R_1 = R_X(D_1) \text{ and } R_1 + \Delta R = R_X(D_2), \tag{2.5}$$

where $R_X(D)$ is the R-D function of the source X at distortion level D. It has been shown in [21] that a necessary and sufficient condition for a source to be successively refinable is that the conditional distributions $f(\hat{X}_1|X)$ and $f(\hat{X}_2|X)$ are Markov compatible in the sense that they can be represented as a Markov chain $X \to \hat{X}_2 \to \hat{X}_1$.

A successive refinement code for the Wyner-Ziv problem consists of multi-stage encoders and decoders where each decoder uses all the information generated from decoders of its earlier stages [4]. Fig. 7 depicts a special case of two-stage successive coding for the Wyner-Ziv problem with the side information at each stage being the same.

Let Y be the side information available to the decoder at both the coarse and the refinement stages, and the corresponding coding rates (distortions) are $R_1(D_1)$ and

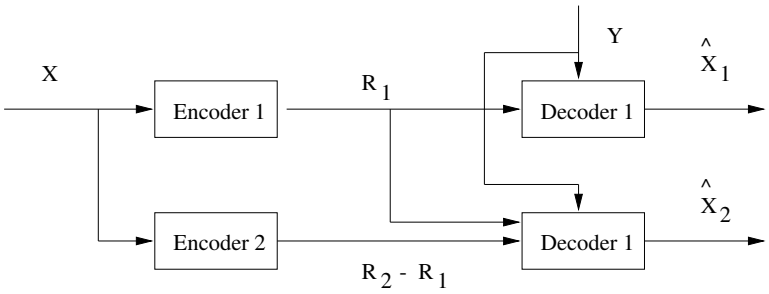


Fig. 7. Two-stage successive refinement with identical side information at the decoders.

$R_2(D_2)$, respectively. Let $R^*_{X|Y}(D)$ be the Wyner-Ziv R-D function [11]. According to (2.5), a source X is said to be *successively refinable* from D_1 to D_2 ($D_1 > D_2$) with side information Y if

$$R_1 = R^*_{X|Y}(D_1) \text{ and } R_1 + \Delta R = R^*_{X|Y}(D_2). \tag{2.6}$$

The notion of successive coding can be naturally extended to any finite number of stages [4]. Consider the case when the side information fed into the K decoders at each level is the same, the source X is *multi-stage successively refinable* with side information Y if

$$R_1 = R^*_{X|Y}(D_1) \text{ and } R_i + \Delta R_i = R^*_{X|Y}(D_{i+1}), \text{ for } i = 1, 2, \ldots k-1. \tag{2.7}$$

Necessary and sufficient conditions for successive refinability are given in [4] and the jointly Gaussian source (with MSE measure) shown to be multi-stage successively refinable in the Wyner-Ziv setting.

F. Successive Wyner-Ziv Coding: Code Design for Ideal Sources

Successive or scalable image/video coding made popular by EZW [42] and 3-D SPIHT [43][44] is attractive in practical applications such as networked multimedia. By producing a video stream which can be decoded at more than one quality level, scalable

video coding achieves graceful degradation of the picture quality, depending on the available bandwidth for data transmission. Hence it is a desirable property for video streaming applications such as video-on-demand. Further, WZC is considered as the enabling technology for asymmetric video coding with simple encoding and relatively complex decoding for "uplink" applications such as sensor networks. Therefore, it is important and rewarding to explore practical code designs on successive WZC.

Extending the successive refinement result of Steinberg and Merhav [4] on joint Gaussian source, Cheng *et al.* [5] proved that the joint Gaussian condition can be relaxed to the case that only the difference between the side information and the source is Gaussian and independent of the side information. Practical layered Wyner-Ziv code design for Gaussian sources based on nested scalar quantization (NSQ) and multi-level LDPC code for SWC was also presented in [5]. Their results are approximately 1.65 to 2.9 dB from the Wyner-ziv bound for rate from 0.48 to 6 bits per sample.

CHAPTER III

LAYERED WYNER-ZIV VIDEO CODING FOR NOISELESS CHANNEL

Although some previous works have been done on practical code designs for ideal sources on SWC and WZC, it is not so straightforward to apply these design techniques directly to video sources. It is intuitive to use the standard decoded video as the side information [1][2][3], which is highly correlated with the original video source. Orthonormal transform should be applied to the video source and the side information to make them conditionally independent before performing WZC, which can be thought of as a quantizer followed by SWC,. However, there are several issues involved in WZC of video sources:

- Transform design: Unlike the ideal sources which are i.i.d. random variables, the neighboring symbols of the video source are highly correlated with each other. In conventional non-distributed source coding, orthonormal transforms are widely used to decorrelate the source vector to facilitate compression. Therefore, for Wyner-Ziv video coding, orthonormal transforms should also be applied to both the source and the side information to make the source symbols conditionally independent given the side information before performing WZC.

- Statistical modeling: In WZC of ideal sources [5], both the source and the noise are assumed to be Gaussian distributed and the Gaussian noise are assumed to be independent of the source. So the problem of Wyner-Ziv video coding involves the correlation modeling of the transformed coefficients of the video source and the side information.

- Quantizer design: The quantizer design for WZC of ideal sources was presented in [41], which assumed the Gaussianality of both the source and the side infor-

mation.

- SWC design and rate control: To approach the Wyner-Ziv limit, capacity-achieving channel code (e.g., turbo [27] and LDPC codes [24]) should be used to approach the Slepian-Wolf limit. However, the performance of these advanced channel codes depends on the long length of the codewords, which can be easily satisfied with ideal sources but incurs long time delay for video sources. In addition, the code rate for SWC and the convergence of the Slepian-Wolf decoding rely on the correlation between the source and the side information.

In this chapter, we will first review recent works by other groups on Wyner-Ziv video coding before presenting our proposed Wyner-Ziv video coding scheme.

A. Previous Works on Wyner-Ziv Video Coding

Recently there have been several works done on applying the WZC principle to video sources:

- Aaron *et al.* [1] proposed a distributed video compression scheme based on turbo codes. Odd frames are coded by the MPEG-4 encoder and even frames coded by a Wyner-Ziv encoder, which consists of scalar quantization and turbo coding. The parity bits of turbo codes are transmitted to the decoder and jointly decoded with the knowledge of the neighboring frames. To exploit the spatial correlation in each frame, they studied transform-domain Wyner-Ziv video coding using the DCT. Simple uniform quantizer were used and the rate control is performed at the Slepian-Wolf decoder by requesting a sufficient number of bits. In [1], they also presented an embedded Wyner-Ziv codec which consists of one base layer and two enhancement layers for graceful video quality

degradation. However, the proposed systems incur a substantial R-D penalty compared to standard MPEG-4 coding. In addition, because parity bits instead of syndrome bits are generated by the Slepian-Wolf encoder, their approach is very similar to systematic coding (the base layer, which is the systematic part, plus the parity bits) but WZC.

- Sehgal *et al.* [2] discussed how coset-based Wyner-Ziv video coding can alleviate the problem of prediction mismatch. Their Wyner-Ziv coding system consists of DCT, dead zone quantization and bit plane coding using regular LDPC codes. However, theoretical quantizer design and SWC design were not explained to justify their approach and no practical applications were presented.

- Puri and Ramchandran [3] outlined a PRISM framework that swaps the encoder/decoder complexity in standard codecs (e.g., MPEG or H.26X). PRISM uses a very simple encoder but a relatively highweight decoder where block-matching motion estimation is performed at the decoder. DCT is applied to each frame, followed by uniform scalar quantizer. Each block is then encoded independently and only the low-frequency coefficients are compressed using syndrome coding based on convolutional codes. Rate control is done at block base depending on estimated statistical dependence. However, only performance at high bit rates was provided in [3].

B. Practical Code Design

We now present our practical successive Wyner-Ziv video coding scheme using LDPC code based bit plane coding for SWC. Treating a standard coded video as the base layer (or side information), a layered Wyner-Ziv bitstream of the original video sequence is generated to enhance the base layer such that it is still decodable with

commensurate qualities at rates corresponding to layer boundaries. Fig. 8 depicts the block diagram of our layered Wyner-Ziv codec, whose encoder consists of three components: the DCT, NSQ [39, 17] and SWC [32, 17] based on irregular LDPC codes [24, 45]. In the first component, we use the DCT as an approximation to the conditional KLT [6][7], which makes the coefficients of the transformed block of the original video X conditionally independent given the same transformed block of the side information Y. NSQ is a binning process that partitions the input DCT coefficients into cosets and outputs only the coset indices. The upper bit planes of the DCT coefficients are skipped in NSQ since they are highly correlated to those in the side information. There will be loss in video quality with this binning process if the side information cannot be used to correctly recover these upper bit planes in the joint Wyner-Ziv decoder. The lower bit planes are less significant and hence quantized to zero by NSQ to save rate. Therefore, both the upper and lower bit planes are thrown away in NSQ and only those in between are coded (see Fig. 9). NSQ introduces both binning loss, which should be kept small with strong *coset/channel* coding, and quantization loss that should be optimally traded off with rate in source coding. In addition, there are still correlation between the quantized version (bit planes in the middle) of the source X and the side information Y [17], and SWC [32] can be employed to exploit this correlation by sending *syndromes* to achieve further compression. We employ multi-level LDPC codes for SWC (or lossless source coding of the quantized source with side information at the decoder) in the third component of the encoder and output one layer of compressed bitstream for each bit plane after NSQ. In doing so, we note that the correlation decreases as we move from MSB to LSB. Thus higher rate LDPC codes are designed for higher bit planes to achieve more compression; while lower rate LDPC codes are given to lower bit planes for less compression. Furthermore, to facilitate layered coding, the order of encoding

proceeds from the MSB to the LSB after NSQ, although theoretically there is no rate difference between the order of bit plane coding,. In the following, we will explain each component in details.

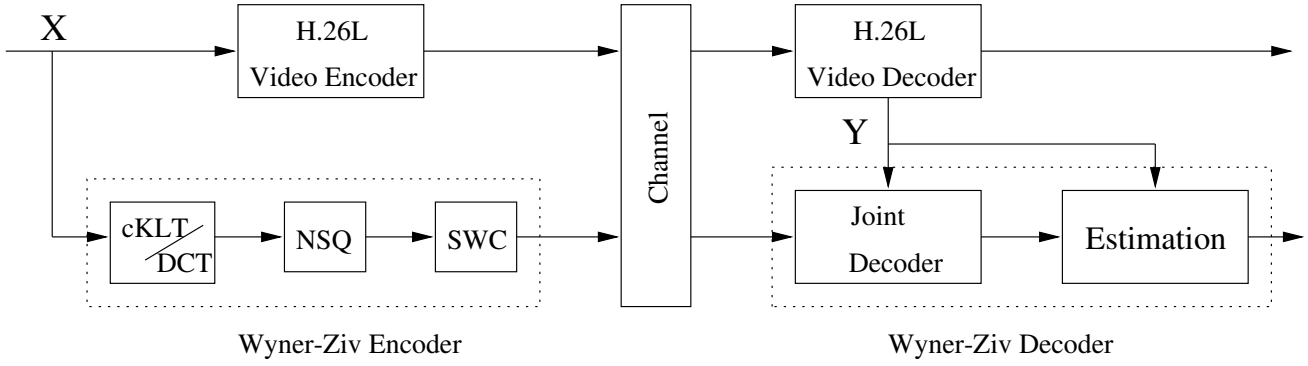


Fig. 8. Block diagram of the proposed layered Wyner-Ziv video codec.

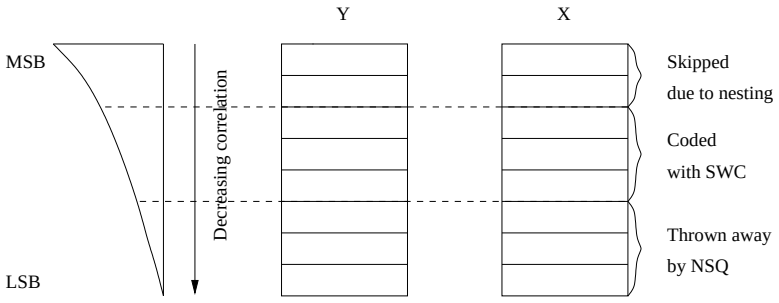


Fig. 9. NSQ throws away both the upper bit planes (with nesting) and the lower bit planes (with quantization).

We denote the current frame of the original video as $\mathbf{x}$ and the H.26L decoded version of $\mathbf{x}$ as $\mathbf{y}$. For Wyner-Ziv coding of $\mathbf{x}$, we first apply the cKLT (approximated by the DCT) to every 4×4 block [7] of $\mathbf{x}$ so that the components of the transformed block $\mathbf{X} = \mathbf{T}\mathbf{x}$ ($\mathbf{T}$ is related to both $\mathbf{x}$ and $\mathbf{y}$) are conditionally independent given the side information $\mathbf{y}$, which is also transformed into $\mathbf{Y} = \mathbf{T}\mathbf{y}$. Each frequency component of $\mathbf{Y}$ (denoted by Y) acts as the side information for the corresponding component of $\mathbf{X}$ (denoted by X). We assume that X and Y are jointly Gaussian with

$X = Y + Z$, where Z is zero-mean Gaussian and independent of X (although DCT coefficients of images/video are better modeled as Laplacian distributed [46]).

The next step is NSQ, which consists of a coarse coset channel code nested in a fine uniform scalar quantizer. Fig. 10 shows a simplest 1-D nested uniform quantizer with $N = 4$ cosets, where the fine source code employs a uniform scalar quantizer with stepsize q and the coarse channel code with minimum distance $d_{min} = Nq$. To encode, X is first quantized by the fine source code (uniform quantizer), resulting an average quantization error of $D_{sc} = q^2/12$ at high rate. However, only the index B $(0 \leq B \leq N-1)$ of the coset in the coarse channel code that the quantized X belongs to is coded to save rate. Using the coded coset index B, the decoder finds in the coset the codeword closest to the side information Y as the best estimate of X. Due to the coset channel code employed in nesting process, the Wyner-Ziv decoder suffers a small probability of error that is inversely proportional to $d_{min} = Nq$. It is desirable to choose a small quantization stepsize q to minimize the distortion D_{sc} due to source coding. On the other hand, d_{min} should be maximized to minimize the distortion D_{cc} due to channel decoding. Thus for a fixed N, there exists an optimal q that minimizes the total distortion $D = D_{sc} + D_{cc}$.

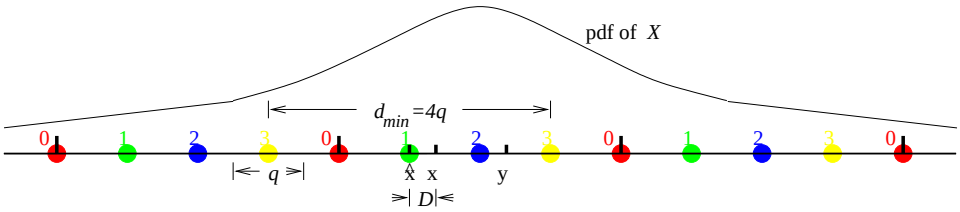


Fig. 10. A 1-D nested scalar quantizer with nesting ratio $N = 4$.

Due to the correlation between X and Y, there still remains correlation between the quantized version B of X and the side information Y. Ideal SWC can be used to compress B to the rate of $R = H(B|Y)$. Express B in its binary representation

as $B = B_0 B_1 \dots B_n$, where B_0 is the MSB and B_n is the LSB. We employ multi-level LDPC codes to compress $B_0 B_1 \dots B_n$ based on the syndrome approach [32, **?**]. The rate of the LDPC code for B_i $(0 \leq i \leq n)$ depends on the conditional entropy $H(B_i|Y, B_{i-1}, \dots, B_0)$ [5], which denotes the minimum rate needed for lossless recovery of B_i given Y and $B_{i-1} \dots B_0$ at the decoder. A specific LDPC code is determined by its bipartite graph, which specifies the connections between the bit nodes and the check nodes. An example of the bipartite graph of a LDPC code is shown in Fig. 11.

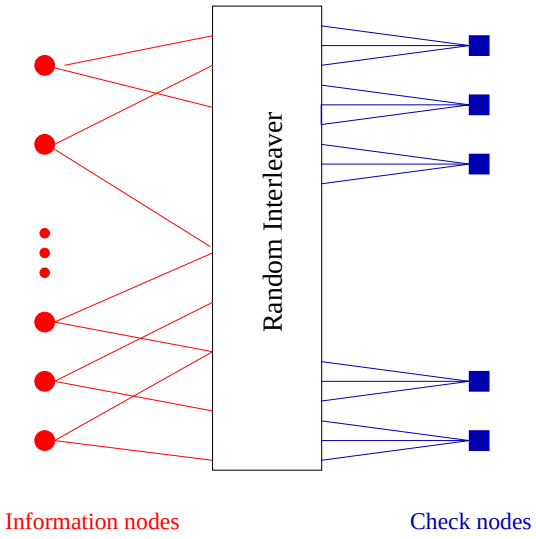


Fig. 11. Graph representation for LDPC codes.

In our simulations, we first assume ideal SWC in the sense that the rate $R = H(B|Y)$ can be achieved. Then for each fixed N (number of cosets in the channel code), we vary the uniform quantization step size q to generate a set of R-D points (R, D) and pick the optimal q^* corresponding to the point with the steepest R-D slope from the zero-rate point in WZC. Note that the distortion for the zero-rate point is just $||X - Y||^2$, which is the average distortion of base layer coding due to H.26L. After identifying the optimal R-D points for different N, the lower convex

hull of these points form the operational R-D curve of WZC. Due to the fact that quadratic Gaussian sources are successively refinable, the same operational R-D curve should be traversed by starting with a large N (with its corresponding q^*) first and then sequentially dropping bit planes of B. In other words, by setting different low bit plane levels of B to zero, the resulting R-D points after Wyner-Ziv decoding should all lie on the operational R-D curve. Our simulations verify this property of successive refinement and justify the practice of coding B_i into the i-th layer with rate $H(B_i|Y, B_{i-1}, \ldots, B_0)$ (see Fig. 12). By the chain rule $H(B|Y) = H(B_0|Y) + H(B_1|B_0, Y) + \ldots + H(B_n|B_0, \ldots B_{n-1}, Y)$. So layered coding suffers no rate loss when compared with monolithic coding.

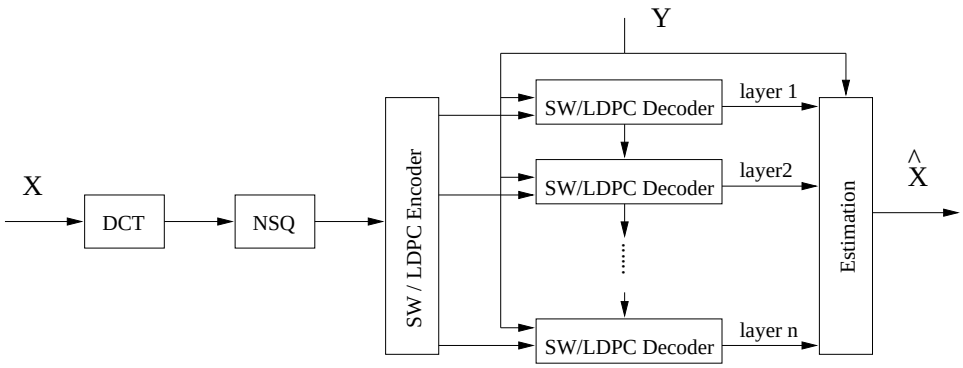


Fig. 12. Bit-plane based multi-stage Slepian-Wolf coding for layered Wyner-Ziv coding after the DCT and NSQ.

In our practical irregular LDPC code design, the code degree distribution polynomials $\lambda(x)$ and $\rho(x)$ of the LDPC codes are optimized using density evolution based on the Gaussian approximation [47]. The bipartite graph (an equivalent representation of the parity-check matrix $\mathbf{H}$) for the irregular LDPC code is then randomly constructed based on the optimized code degree polynomials $\lambda(x)$ and $\rho(x)$. To compress bit plane B_i, only the corresponding syndromes determined by the sparse parity check matrix of the irregular LDPC code are coded. At the decoder, the received syn-

drome bits for each layer (or bit plane) will be combined with the decoded bits of previous bit planes and the side information Y to perform joint decoding. Let $\hat{B}_i$ represent the reconstruction of B_i. The message-passing algorithm [48] is used for iterative LDPC decoding, in which the received syndrome bits correspond to the check nodes on the bipartite graph, the side information and the previously decoded bit planes provide the *a priori* information as to how much is the probability that the current bit is "1" or "0", i.e.,

$$LLR = \log \frac{p(B_i = 0 \,|\, Y, \hat{B}_0, \ldots, \hat{B}_{i-1})}{p(B_i = 1 \,|\, Y, \hat{B}_0, \ldots, \hat{B}_{i-1})}. \tag{3.1}$$

At the decoder, each additional bitstream/syndrome layer is combined with previously decoded bit planes to decoder a new bit plane before joint estimation of the output video. After decoding B_0 as $\hat{B}_0$, both $\hat{B}_0$ and Y will be fed into the decoder for the decoding of B_1. Since the allocated bit rate for coding B_1 is $H(B_1|Y, B_0)$, B_1 can be correctly decoded as long as $\hat{B}_0 = B_0$. By multi-stage decoding, B_i can be correctly recovered with the help of Y and the previously decoded bit planes $B_0, B_1, \cdots, B_{i-1}$, which are already available at the decoder. The more syndrome layers the decoder receives or the higher the bit rate, the more bit planes of B will be recovered to better reconstruct X. Therefore, successive WZC provides the flexibility to accommodate a wide range of bit rates. Progressive decoding is desirable for applications where only a coarse description of the source suffices at the first stage with low bit rate, and fine details are needed at some later stage with higher bit rate.

We perform optimal estimation at the joint decoder. The decoded coset index $\hat{B}_0\hat{B}_1 \ldots \hat{B}_i$ specifies the uncertainty region of X. The side information essentially supplies the conditional PDF of X given Y, which is that of a Gaussian with mean Y and variance proportional to the correlation between Y and X. The optimal estimate of X is computed as the conditional centroid $\hat{X} = E(X|\hat{B}_0\hat{B}_1 \ldots \hat{B}_i, Y)$. Finally, the

inverse DCT is applied to $\hat{X}$ to obtain $\hat{\mathbf{x}}$ in the pixel domain.

C. Experimental Results of Coding Efficiency

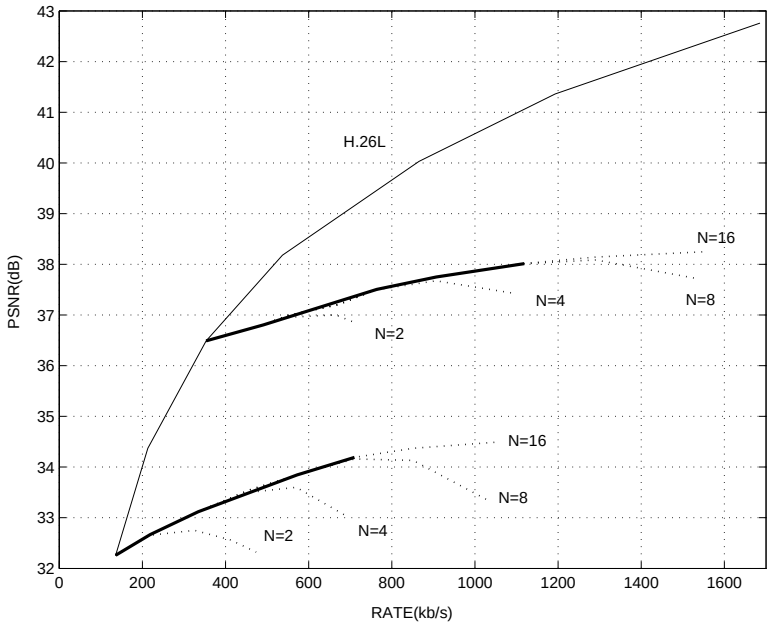
1. Successive Refinement

Due to the approximation of the cKLT by the DCT and the Gaussian assumption of X and Y in our practical Wyner-Ziv video coder, experiments are carried out on the CIF Foreman sequence to verify the validity of our practice and illustrate successive refinement.

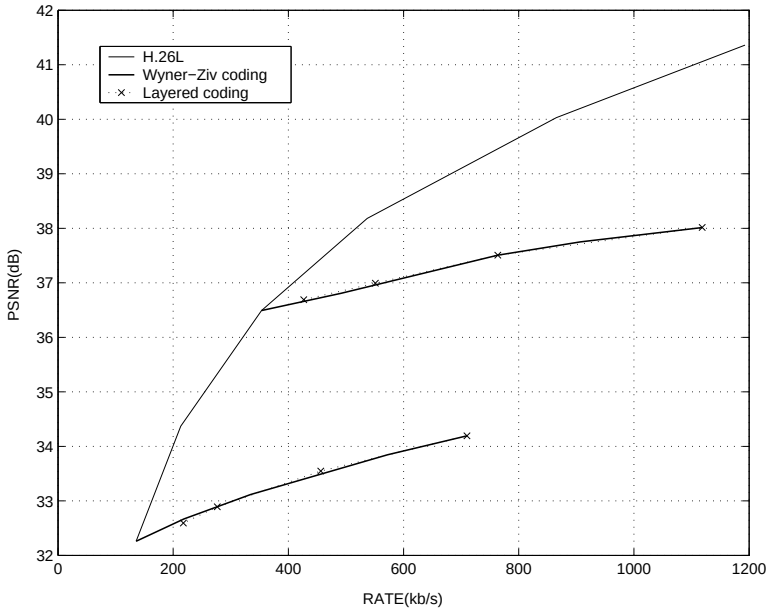
Ideal SWC is assumed in this subsection so that we can use the computed $R = H(B|X)$ as the rate. The R-D performance of four different values of $N \in \{2, 4, 8, 16\}$ with different q's for each N, starting from two different zero-rate points, are plotted in Fig. 13. Fig. 13 (a) illustrates the generation of the operational R-PSNR function of WZC as the upper concave hull of different R-PSNR points. Fig. 13 (b) shows good match between the performance of optimal WZC and that of layered WZC. Layered coding is done by starting at a high rate (e.g., with $N = 16$ and its corresponding q^*) and dropping more and more bit planes of B to achieve lower rates. Thus we have agreement between theory and practice.

2. Layered Coding

We implement SWC based on irregular LDPC codes and investigate the layered WZC performance for CIF sequences "Foreman" and "Mother_daughter". Standard H.26L encoded video is treated as side information at the decoder. One hundred frames are compressed with a frame rate of 30Hz. For each of these sequences, the first frame is coded as I frame, and all the subsequent frames as P frames by H.26L. Different quantization stepsizes are used in the H.26L coder to generate different "zero-rate"



(a)



(b)

Fig. 13. Illustration of successive refinement in our layered Wyner-Ziv video coder, assuming ideal SWC. (a) The operational R-PSNR function of WZC is formed as the upper concave hull of different R-PSNR points. (b) There is almost no performance loss between optimal WZC and layered WZC.

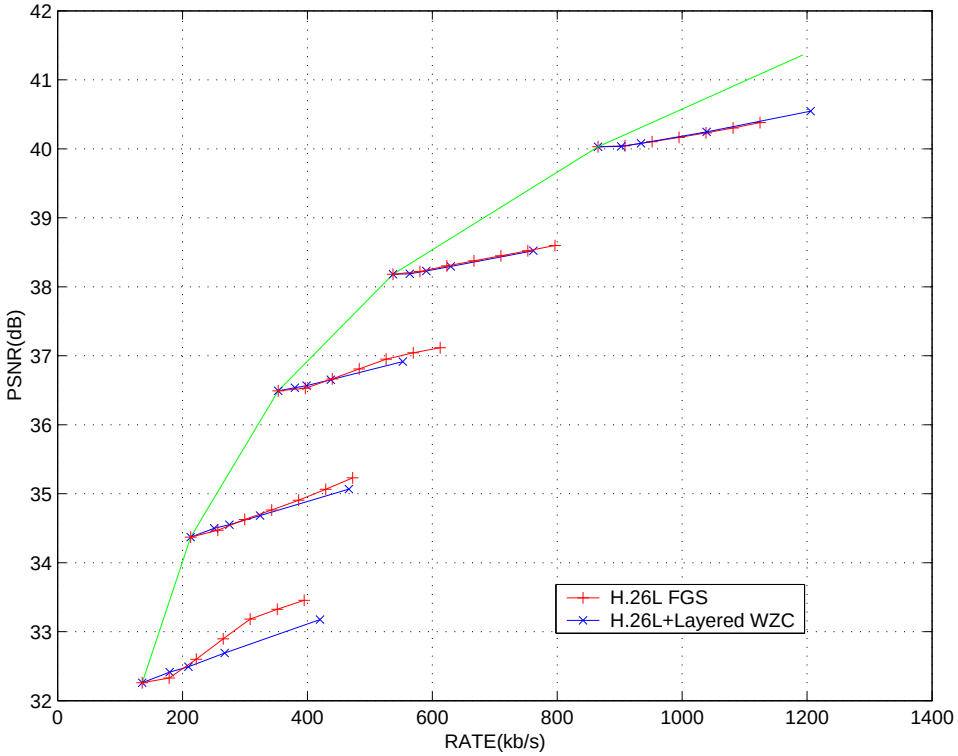


Fig. 14. Layered WZC of the CIF Foreman sequences, starting from different "zero-rate" points. The sum of the rates for H.26L coding and WZC is shown in the horizontal axis.

points for WZC.

In WZC, we assume that $X = Y + Z$ in the DCT domain, where the side information $Y \sim N(0, \sigma_Y^2)$ and the quantization noise $Z \sim N(0, \sigma_Z^2)$ due to H.26L coding are independent. We estimate σ_Z^2 based on the quality of the H.26L decoded sequence (i.e., the side information Y). After NSQ of DCT coefficients of X, irregular LDPC codes with different code rates are used for different bit planes. The LDPC code rate for the i-th bit plane B_i is maximized to achieve the conditional entropy $H(B_i|Y, B_{i-1}, \ldots, B_0)$. The profiles of the LDPC codes are optimized using Gaussian approximation [47] and the structure of them were generated randomly. The block length of each LDPC code is 10^5, which require the grouping of 20 frames to be coded

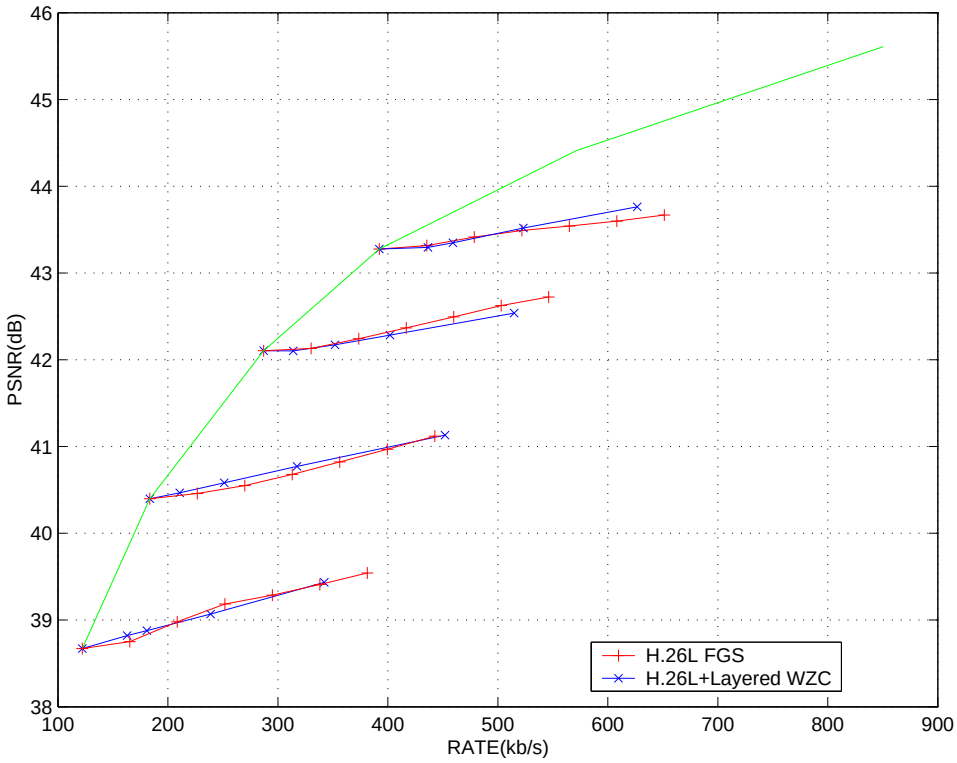


Fig. 15. Layered WZC of the CIF Mother_daughter sequences, starting from different "zero-rate" points. The sum of the rates for H.26L coding and WZC is shown in the horizontal axis.

together. The same pseudo-random seed is used at both the encoder and the decoder such that the codebooks used are the same. The joint decoder performs optimal estimation based on the side information Y and the decoded coset index. Compared to ideal SWC with $R = H(B|X)$, the loss due to practical LDPC coding is about 0.3 dB in PSNR.

Starting with the largest $N = 16$ and its corresponding optimal q^*, we quantize X into B and sequentially decode B_0, B_1, B_3 and B_4. Layered WZC results in terms of rate vs. average PSNR is shown in Fig. 14 for "Foreman" and in Fig. 15 for "Mother_daughter". We see that as more bit planes are decoded, the video quality improves. The overall loss due to layered Wyner-Ziv coding over H.26L monolithic

coding is 1.5 to 4 dB. We also observe that the performance loss due to WZC from H.26L monolithic coding is less when the bit rate for base layer H.26L coding (or "zero-rate" for WZC) is higher. This is partially because the correlation between X and Y is higher when the base layer is coded at higher rate with better quality.

For the first 20 frames of "Foreman" sequence with the starting point at about 530 Kbps, the theoretical rate limit and the actual LDPC code rate used for each bit plane after the NSQ of the DC component and the first two AC components of the DCT coefficients are listed in Table II in Appendix A. The quantization stepsize for the NSQ is $q = 32$, and the LDPC code length of each bit plane is 112160.

The designed degree profiles of the LDPC codes with code rate from $0.50 \sim 0.90$ are listed in Table III from (a) to (d) in Appendix A.

3. Wyner-Ziv Coding for Error Robustness

Our layered Wyner-Ziv video coding framework is very similar to FGS coding [19][20] in the sense that both schemes treat the standard coded video as the base layer and generate an embedded bitstream as the enhancement layer. However, the key difference is that instead of coding the difference between the original video and the base layer reconstruction like FGS, the enhancement layer is generated "blindly" without knowing the base layer in Wyner-Ziv video coding. Therefore, the stringent requirement of FGS coding that the base layer is always available losslessly at the decoder/receiver can be lossened somewhat as an error-concealed version of the base layer can still be used in the joint Wyner-Ziv decoder.

In our experiment, we compressed the CIF sequence "Foreman" using both the Wyner-Ziv video coder and the H.26L-FGS [20] coder at a frame rate of 30Hz. The base layer is encoded at about 190 Kb/s and the bit rate for the enhancement layer of both the Wyner-Ziv coding and the FGS coding is about 60 Kb/s. One intra

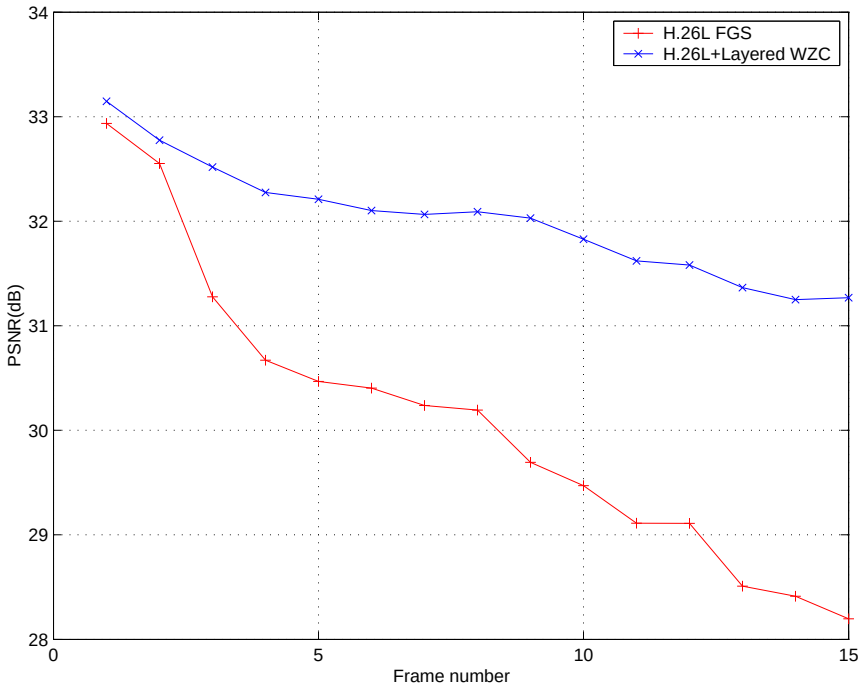


Fig. 16. Error resilience performance of Wyner-Ziv video coding compared with H.26L-FGS.

frame (I frame) is inserted every 15 frames and the rest frames are all P frames. Then 1% macroblock loss at the base layer is simulated. Simple error concealment is performed during the decoding for base layer. The coding efficiency of the first 15 frames is shown in Fig. 16. The performance of Wyner-Ziv video coding is about 2 dB better on average than H.26L-FGS coding in case of base layer packet losses. This is because the basic assumption of FGS coding is no longer valid in this setup while the error-concealed version of the base layer can still be used as side information for decoding in Wyner-Ziv video coding system.

As an example, the decoded $10th$ frames of CIF "foreman" sequence by H.26L-FGS and WZC in the previous simulation are shown in Fig. 17 (a) and (b) respectively. Obviously, the decoded video in Fig. 17 (b) has higher visual quality than that in Fig. 17 (a). Therefore, Wyner-Ziv video coding exhibits inherent robustness

(a)

(b)

Fig. 17. Substantial improvement in decoded video quality is observed by using Wyner-Ziv video coding scheme. (a) The $10th$ decoded frame by H.26L-FGS. (b) The $10th$ decoded frame by Wyner-Ziv video coding.

in case of channel errors in the base layer.

D. Open Issues

To summarize, our proposed system solves the problems mentioned in the beginning of the chapter in the following way:

- The low-complexity DCT is used as an approximation to the conditional KLT.
- The DCT coefficients of the source and the side information are both modeled as Gaussian distributions and their variances are estimated by collecting statistics.
- We select the optimal nesting ratio N and the stepsize q of the NSQ by generating the operational R-PSNR points assuming ideal SWC.
- The rate of the SWC for each bit plane depends on the corresponding conditional entropy given the side information and all the previous bit planes, which is the theoretical rate limit for successful recovery.

In addition, our system provides the scalability of video coding in the sense that the layered Wyner-Ziv bitstream enhances the base layer such that it is still decodable with commensurate qualities at rate corresponding to layer boundaries.

However, there are still some open issues:

- More accurate statistical modeling will definitely improve the NSQ and the SWC design at the encoder and the Slepian-Wolf decoding and the optimal estimation at the decoder.
- LDPC code design can be further improved using density evolution without Gaussian approximation for different source distributions. In addition, it is

desirable to improve the performance of LDPC codes with shorter block length to reduce the time delay in video coding.

- Our current implementation of layered Wyner-Ziv coding only allows decoding at layer boundaries decoding at the middle of a layer (bit plane) will suffer a huge performance loss as unavailable bits in the layer have to be treated as erasures. This is due to the limitation of the LDPC codes used for SWC. Ideally we would prefer progressively decodable channel (or Slepian-Wolf) codes, much like arithmetic codes used in classic source coding.

CHAPTER IV

LAYERED WYNER-ZIV VIDEO CODING FOR NOISY CHANNELS

In the previous chapter, we presented a practical layered Wyner-Ziv video coding system using LDPC code based bit plane coding for SWC. In this chapter, we remove the assumption of error free channel and propose a successive Wyner-Ziv video coding scheme for BSC using IRA code [22][23] based bit plane coding for SWC.

A. Problem Formulation

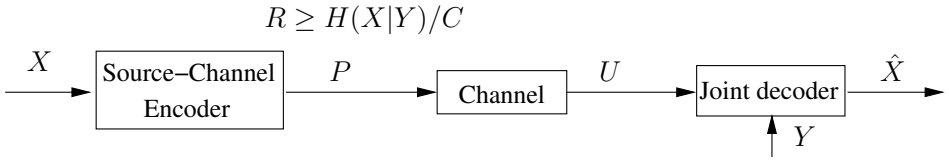


Fig. 18. Lossless JSCC of X with side information Y at the decoder.

Consider two i.i.d. sequence $\{X_i\}_{i=1}^{\infty}$ and $\{Y_i\}_{i=1}^{\infty}$ which are correlated by a given distribution $p_{XY}(x, y)$. A special case of the DSC of X and Y over noisy channels is the lossless JSCC of the source with side information at the decoder, i.e., when Y is losslessly available at the joint decoder and we try to compress X as efficiently as possible and still be able to recover X correctly after transmitting the coded version through the noisy channel, as illustrated in Fig. 18. According to the Slepian-Wolf theorem [10] for DSC over noiseless channel, there is no loss of coding efficiency with separate encoding when compared to joint encoding as long as joint decoding is performed. Assume Y is encoded at its entropy rate $H(Y)$ such that it is losslessly available at the joint decoder, then the theoretical limit for lossless coding of X over an ideal channel is the conditional entropy $H(X|Y)$. However, for nonideal channel

with capacity C, the theoretical limit for lossless coding of X becomes $H(X|Y)/C$ due to the separation principle [51]. Therefore, the discrete noisy channel incurs an extra rate loss of $1/C$, where $C = 1 - H(q) = 1 + q\log_2 q + (1-q)\log_2(1-q)$ for the BSC with COP q.

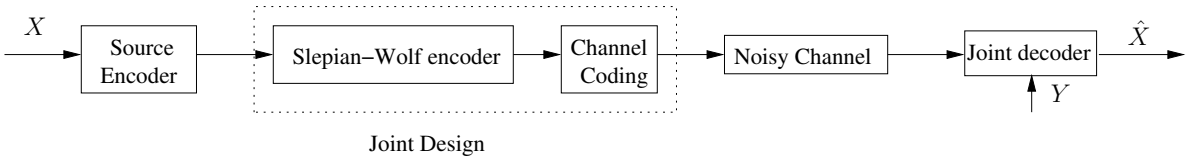


Fig. 19. Lossy JSCC of X with side information at the decoder over noisy channels.

For the more general case of lossy JSCC of X with side information at the decoder, since the achievable lower bound for the bit-rate for an expected distortion D for WZC of X is $R^*_{WZ}(D)$ [11], the theoretical R-D function becomes $R(D) = R^*_{WZ}(D)/C$ due to the separation principle [51]. To achieve this theoretical rate limit, one can employ WZC followed by additional channel coding for noisy channels (Fig. 19). WZC can be thought of as a quantizer followed by SWC, which is a channel coding problem. If the channel is noisy, extra loss is introduced due to its channel capacity and additional channel coding component is required to provide error protection. The channel coding required can be combined with the SWC component of WZC in a joint design as a stronger channel code. Capacity-approaching channel codes such as LDPC codes are used to achieve the Slepian-Wolf limit for error free channel. For noisy channels, it is straightforward to add an additional channel code to the syndrome bits of the LDPC codes. IRA codes, which is a special class of LDPC codes, can be used for SWC over noisy channels.

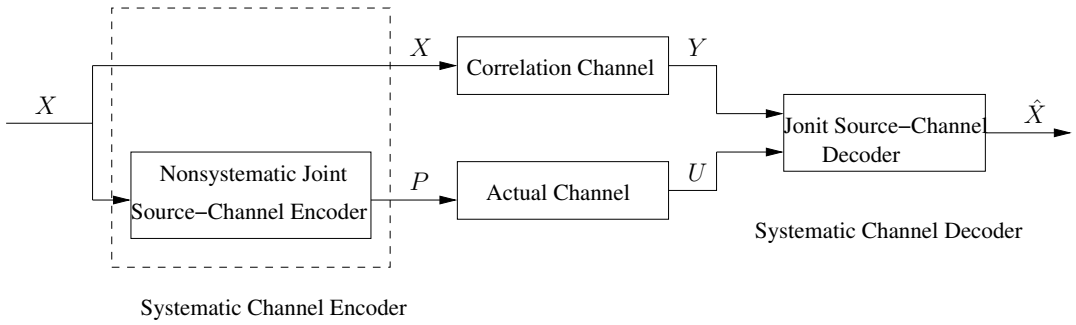


Fig. 20. The two equivalent channels for JSCC of X with side information Y at the decoder.

B. Previous Work

JSCC with side information at the decoder using IRA codes based on the idea of two equivalent channels was done in [23] for binary sources. Liveries *et al.* considered the problem of JSCC with side information at the decoder for a nonequiprobable memoryless binary source, where the binary source X_i and the side information Y_i are correlated with $P(X_i \neq Y_i) = p$. Therefore, the theoretical rate limit for lossless recover of X becomes $H(p)/C$ since $H(X|Y) = H(p)$ (C is the capacity of the noisy channel). Their approach of code design is based on viewing the correlation between the binary source output and the side information as a separate channel, i.e., X_i will be the input to a BSC with COP p and Y_i its distorted output. Only the parity bits P, which is the coded version of X, are transmitted through the actual noisy channel. Then the alternative way of viewing Fig. 18 is the transmission of a single code word over two different channels [23], as illustrated in Fig. 20. The systematic part X goes through the correlation channel and the parity part P through the actual channel. Systematic IRA codes are used as the source-channel codes and designed using the Gaussian approximation approach [22] by taking into account

the two different channel conditions. Their simulated performance results are better than other solutions using turbo codes over the BSC, AWGN and flat Rayleigh fading channel.

C. Wyner-Ziv Video Coding using IRA Codes for the BSC

In this thesis, we apply the JSCC techniques with IRA codes to video sources. Fig. 21 depicts the block diagram of the proposed layered Wyner-Ziv video codec using IRA code based bit plane coding for SWC. The standard coded video, which can be viewed as the output of the correlation channel, is the received systematic part of the IRA code and treated as the side information at the decoder. The Wyner-Ziv coded bitstream can be viewed as the parity part of the IRA code and is transmitted through the BSC. At the decoder, the distorted Wyner-Ziv coded bits are combined with the side information to perform joint source-channel decoding. In other words, both the systematic part and the parity part of the IRA codes are used to perform systematic channel decoding.

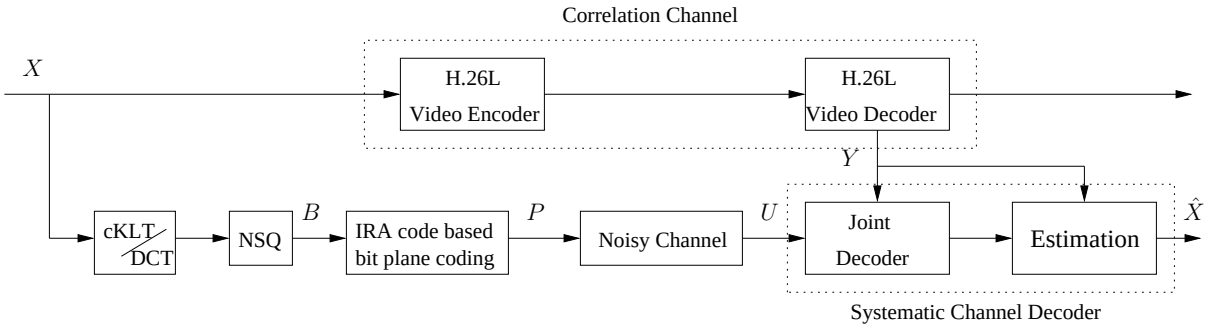


Fig. 21. Block diagram of Wyner- Ziv video coding using IRA codes for noisy channels.

The Wyner-Ziv encoder consists of three components: the DCT, NSQ [39, 17] and SWC [32, 17] based on IRA codes [22][23]. The DCT and NSQ play the same

roles as before. We employ multi-level IRA codes for SWC in the third component of the encoder and output one layer of parity bits for each bit plane after NSQ. The IRA codes based bit plane coding for SWC not only exploits the correlation between the quantized version of the source X and the side information Y, but also provides error protection for the parity bits that are transmitted through the BSC. At the decoder, the received parity bits of each layer are combined with previously decoded bit planes to decode a new bit plane before joint estimation of the output video. This multi-level decoding scheme achieves progressive decoding because the decoded video quality gradually improves upon receiving more layers of parity bits.

Since the proposed system in this chapter differs from that in the previous chapter only in the Slepian-Wolf coder, we will constrain ourselves to the discussion of IRA code based bit plane coding for SWC.

IRA codes were introduced in [22] and it was shown that IRA codes can achieve capacity on the erasure channel and can perform very close to the capacity on the binary input WGN channel. Systematic IRA codes have the advantages of both LDPC codes (message-passing iterative decoding) and turbo codes (linear time encoding). An emsemble of IRA code is described by the degree distribution polynomials $\lambda(x) = \Sigma_{i=2}^{J} \lambda_i x^{i-1}$ and $\rho(x) = \Sigma_{i=2}^{M} \rho_i x^{i-1}$, where λ_i and ρ_i are the fraction of edges incident on information nodes and check nodes with degree i, respectively. In this chapter, the check degrees are assumed to be concentrated, i.e., $\rho(x) = x^{\alpha-1}$ for some integer α. A specific IRA code is determined by its bipartite graph, which specifies the connections between the bit nodes and the check nodes. An example of the bipartite graph of a IRA code is shown in Fig. 22.

We only consider IRA codes in their systematic form in this thesis, that is, the codeword corresponding to information bits $\{b_1, \ldots, b_k\}$ is given by $\{b_1, \ldots, b_k, p_1, \ldots, p_r\}$, where $\{p_1, \ldots, p_r\}$ are the parity bits. Given the degree distribution polynomial $\lambda(x)$

and α, The rate of the code is:

$$R = \frac{\Sigma_i \lambda_i / i}{1/\alpha + \Sigma_i \lambda_i / i}.$$

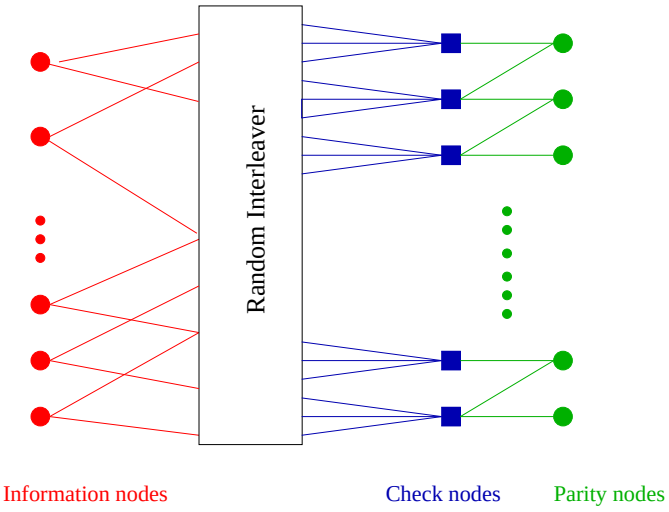


Fig. 22. Graph representation for IRA codes.

1. Slepian-Wolf Encoding

Due to the correlation between X and Y, there still remains correlation between the quantized version B of X and the side information Y (Fig. 21). If the channel is error free, ideal SWC can be used to compress B to the rate of $R = H(B|Y)$. However, the actual noisy channel incur an extra rate loss of $1/C$ (C is the capacity of the noisy channel), and additional channel coding component is required to provide error protection. IRA codes are employed for SWC over noisy channels to achieve the theoretical rate limit of $R = H(B|Y)/C$.

Express B in its binary representation as $B = B_0 B_1 \dots B_n$, where B_0 is the MSB and B_n is the LSB. We employ multi-level IRA codes to generate the parity bits of $B_0 B_1 \dots B_n$. The rate of the IRA code for B_i ($0 \leq i \leq n$) depends on $H(B_i|Y, B_{i-1}, \dots, B_0)/C$ [23], which denotes the minimum rate needed for lossless

recovery of B_i given Y and $B_{i-1} \dots B_0$ at the decoder over the BSC with capacity C. By the chain rule $H(B|Y)/C = H(B_0|Y)/C + H(B_1|B_0, Y)/C + \dots + H(B_n|B_0, \dots B_{n-1}, Y)/C$. So layered coding suffers no rate loss when compared with monolithic coding.

To compress bit plane B_i, only the corresponding parity bits P_i determined by the code structure of the IRA code are coded. Given a realization of the bipartite graph with degree distribution $\lambda(x)$ and $\rho(x) = x^{\alpha-1}$, the encoding of an input information sequence of length k is performed by mapping the bit nodes to the parity nodes, generating a binary sequence of length $r = \frac{k}{(\alpha \Sigma_i \lambda_i / i)}$.

2. Joint Decoding

At the decoder, the received parity bits U from the BSC and the corresponding systematic bits which can be generated from the side information are combined to perform systematic channel decoding. The message-passing algorithm [22] is used for iterative IRA decoding, in which the received syndrome bits correspond to the parity nodes on the bipartite graph, the side information and the previously decoded bit planes provide the *a priori* information as to how much is the probability that the current bit is "1" or "0". Therefore, the LLR of the jth systematic node at the ith layer is

$$LLR_{sys}^{(i,j)} = \log \frac{p\,(B_{i,j} = 0\,|\,Y, \hat{B}_0, \dots, \hat{B}_{i-1})}{p\,(B_{i,j} = 1\,|\,Y, \hat{B}_0, \dots, \hat{B}_{i-1})}. \tag{4.1}$$

The LLR of the jth parity node at the ith layer depends on the received parity bits $u_{i,j}$ and the actual channel condition. For the BSC with COP q,

$$\begin{aligned} LLR_{par}^{(i,j)} &= \log \frac{p\,(p_{i,j} = 0|u_{i,j})}{p\,(p_{i,j} = 1|u_{i,j})} \\ &= (1 - 2u_{i,j}) \log \frac{1-q}{q}. \end{aligned} \tag{4.2}$$

Then the message-passing algorithm [22] is used for iterative IRA decoding.

After the decoding of the ith layer, both the decoded bit plane and the side information Y will be fed into the decoder for the decoding of subsequent bit planes. Since the allocated bit rate for coding B_i is $H(B_i|Y, B_0, \ldots, B_{i-1})/C$, B_i can be correctly decoded as long as B_{i-1} is correctly recovered. The more syndrome layers the decoder receives or the higher the bit rate, the more bit planes of B will be recovered to better reconstruct X. Therefore, successive Wyner-Ziv coding provides the flexibility to accommodate a wide range of bit rates over the noisy channels.

3. Code Design

The possibility of designing systematic IRA codes with different channel conditions for the systematic and the parity part is the main advantage of using IRA codes in JSCC with side information. For a single Gaussian channel, the code degree distribution polynomials of the systematic IRA codes can be optimized using Gaussian approximation. We start by defining the function

$$\phi(t) = \frac{1}{\sqrt{4\pi t}} \int_{-\infty}^{\infty} \tanh(\frac{\alpha}{2}) e^{-(a-t)^2/4t} da. \tag{4.3}$$

Assuming that the maximum allowable systematic node degree is J, the linear optimization of the systematic node degree distribution $\lambda(x) = \Sigma_{i=2}^{J} \lambda_i x^{i-1}$ for a given check node degree distribution $\rho(x) = x^{\alpha-1}$ is done by maximizing $\Sigma_{i=2}^{J} \frac{\lambda_i}{i}$ subject to the condition [22]

$$\lambda(1) = 1, \tag{4.4}$$

$$F(x) > x, \quad \forall x \in [x_0, 1], \tag{4.5}$$

where $F(x)$ is defined as [22]

$$F(x) = \Sigma_{i=1}^{J} \lambda_i \phi(\mu_{sys} + (i-1)\phi^{-1}(\frac{\phi^2(f(x))}{x^{\alpha+1}})), \tag{4.6}$$

and $x_0 = \phi(\mu_{sys})$. The function $f(x)$ is determined from the equation $\phi(f(x)) = x^{\alpha}\phi(\mu_{par} + f(x))$ [22]. The systematic part and the parity part have the same channel LLR $\mu_{sys} = \mu_{par} = 4\frac{E_s}{N_0}$ since they are transmitted through the same Gaussian channel.

To apply this design process using Gaussian approximation to JSCC over two different channels, the only parameters we need to define in the above linear optimization procedure are the initial LLRs obtained from the channels for the systematic part and the parity part, the former from the correlation channel and the latter from the actual channel. As the BSC can be related to AWGN channel as one with quantized output, an approximate way to design the code [23] is to first mapping the BSC channel parameters to that of the AWGN and then using the process outlined above. The mapping is based on the equality of the stability functions for the two channels [23][52]

$$(\frac{E_s}{N_0})_{\text{eq}} = -\log(2\sqrt{q(1-q)}), \tag{4.7}$$

where q is the BSC COP and $(\frac{E_s}{N_0})_{\text{eq}}$ is the equivalent AWGN channel parameter.

D. Experiment Results

Wyner-Ziv video coding system using IRA code based bit plane coding for SWC is implemented and the performance for CIF sequences "Foreman" and "Mother_daughter" is investigated. Standard H.26L encoded video is treated as side information at the decoder. One hundred frames are compressed with a frame rate of 30Hz. For each of these sequences, the first frame is coded as I frame, and all the subsequent frames as

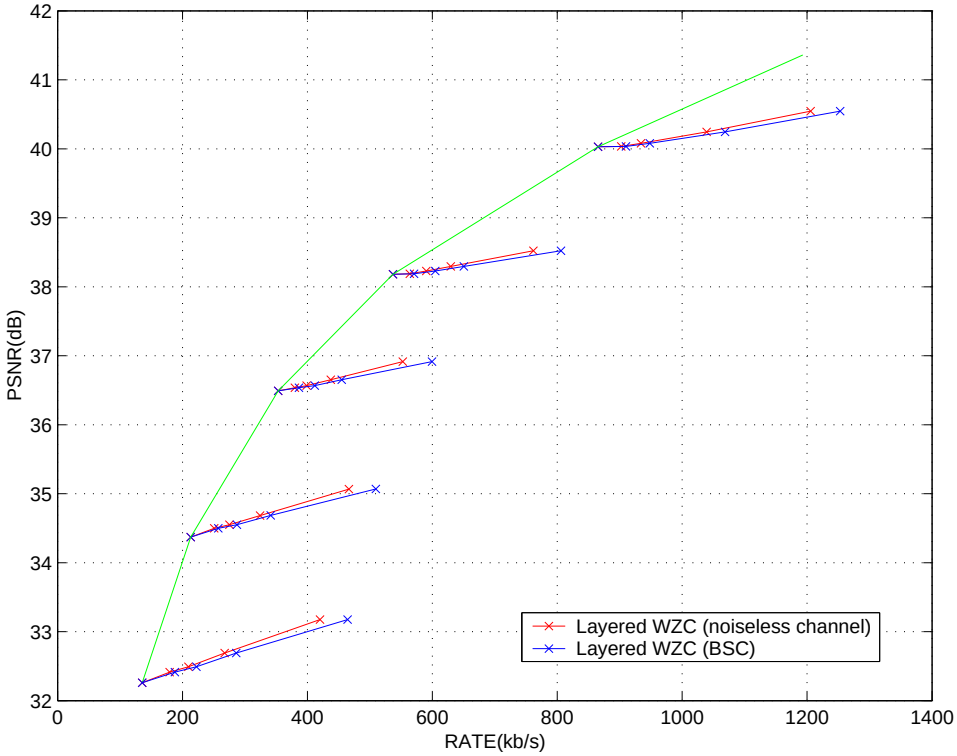


Fig. 23. Performance of Wyner-Ziv video coding with IRA codes over the BSC with COP 0.01 for Foreman sequence.

P frames by H.26L. Different quantization stepsizes are used in the H.26L coder to generate different "zero-rate" points for WZC.

In WZC, we assume that $X = Y + Z$ in the DCT domain, where the side information $Y \sim N(0, \sigma_Y^2)$ and the quantization noise $Z \sim N(0, \sigma_Z^2)$ due to H.26L coding are independent. We estimate σ_Z^2 based on the quality of the H.26L decoded sequence (i.e., the side information Y). After NSQ of DCT coefficients of X, each bit plane is encoded with IRA codes and the parity bits are transmitted through the actual noisy channel. The rate of the IRA code depends on both the conditional entropy $H(B_i|Y, B_{i-1}, \ldots, B_0)$ and the channel capacity C. The profiles of the IRA codes are optimized using Gaussian approximation [22] based on the different channel conditions on the correlation channel and the actual channel. The block length of

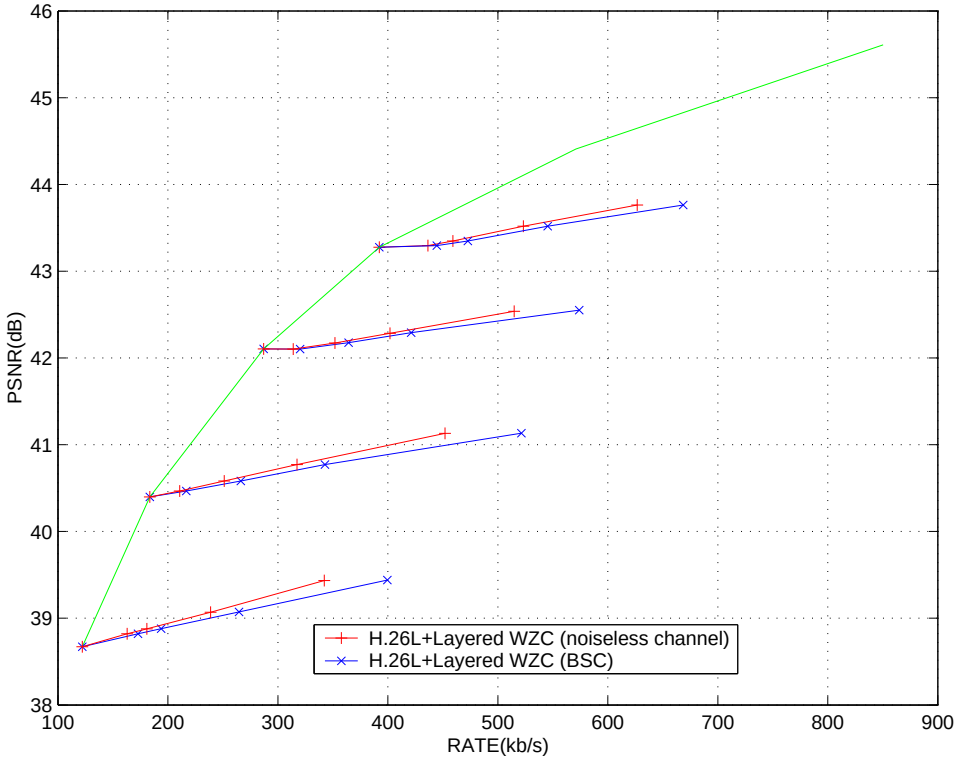


Fig. 24. Performance of Wyner-Ziv video coding with IRA codes over the BSC with COP 0.01 for Mother_daughter sequence.

each IRA code is 10^5, which requires the grouping of 20 frames to be coded together.

Starting with the largest $N = 16$, we quantize X into B and encode each bit plane of B with IRA codes at corresponding code rates. At the decoder, the received parity bits from the actual channel and the side information are combined to perform sequential decoding. Layered Wyner-Ziv video coding results in terms of rate vs. average PSNR is shown in Fig. 23 for "Foreman" and in Fig. 24 for "Mother_daughter". We see that as the bit rate increases and more bit planes are received, the decoded video quality gradually improves. Compared to ideal SWC with $R = H(B|Y)/C$, the loss due to practical SWC using IRA codes is only about 0.08 b/s.

For the first 20 frames of "Foreman" sequence with the starting point at about 530 Kbps, the theoretical rate limit and the actual IRA code rate used for each bit

Table I. The theoretical rate limit and the actual IRA code rate used for each bit plane after the NSQ of the DC component (a) and the first two AC components (b, c) of the DCT coefficients.

ith bit plane	$H(B_i\|Y, B_{i-1}, \ldots, B_0)/C$	IRA code rate
0	0.03734	0.93
1	0.04226	0.93
2	0.11050	0.85
3	0.21315	0.77

(a)

ith bit plane	$H(B_i\|Y, B_{i-1}, \ldots, B_0)/C$	IRA code rate
0	0.00193	0.98
1	0.04099	0.93
2	0.02755	0.94
3	0.18226	0.78

(b)

ith bit plane	$H(B_i\|Y, B_{i-1}, \ldots, B_0)/C$	IRA code rate
0	0.03030	0.93
1	0.00531	0.98
2	0.03525	0.93
3	0.48184	0.56

(c)

plane after the NSQ of the DC component and the first two AC components of the DCT coefficients are listed in Table I. The quantization stepsize for the NSQ is $q = 32$, and the length of the systematic bits of the IRA code for each bit plane is 112160.

CHAPTER V

LAYERED WYNER-ZIV VIDEO CODING FOR PACKET ERASURE CHANNEL

In this chapter, we consider layered Wyner-Ziv video coding with UEP for packet erasure channel. By using UEP, we try to find the optimal source-channel coding trade-off to provide error robustness for transmission of Wyner-Ziv coded bitstreams in fixed-length packets over packet erasure channel.

A. Unequal Error Protection

Besides efficient compression, scalability is a desirable property for video streaming applications such as video-on-demand. A scalable bit stream can be decoded at different rates with commensurate reconstruction quality [42][43][44]. Therefore, the source needs to be encoded only once for different quality requirements. Moreover, scalable encoding enables progressive transmission and bandwidth adaptation. In image transmission, for example, a receiver does not have to wait until all bits are received before decoding the image, instead it can use additional received bits to improve the quality of the previous reconstructed image.

Recently, scalable image and video transmission over various channels have been considered in [55][56][58][60][61][62][63]. These schemes generally involve embedded source coding [43][44], channel coding [53] and JSCC [55][61]. The objective of optimal UEP for multimedia data transmission is to minimize the average MSE under a transmission rate constraint and known channel condition. Because of the huge number of all candidate solutions, such an optimization problem is generally very time-consuming. Exhaustive search is thus prohibitive and various algorithms have been developed. In [55], Chande *et al.* addressed the problem of JSCC of images for progressive transmission over memoryless bit error or packet erasure channels with

fixed source block size and variable channel codeword length. For a target transmission rate, they devised an dynamic programming algorithm to obtain an assignment of the available channel codes to the information blocks by considering three performance measures: the average distortion, the average PSNR and the average useful source coding rate. In [62], UEP was used to provide more efficient error control over equal error protection (EEP). A fast UEP algorithms for real-time transmission of fixed-length packets of the embedded multimedia data is designed with systematic Reed-Solomon (RS) codes for packet erasure channels and with product code for fading channels. In this chapter, we will apply the UEP scheme proposed in [62] to the embedded bitstream generated by successive Wyner-Ziv video coding for transmission over packet erasure channel.

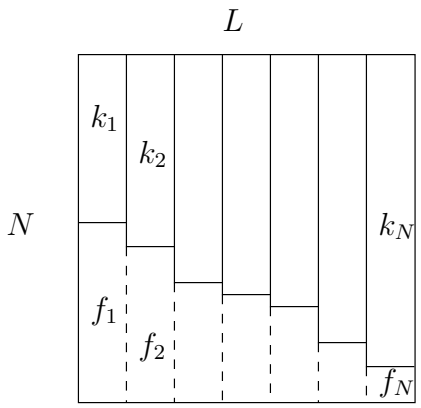


Fig. 25. UEP using RS codes. There are N packets of L symbols each.

Consider a system that transforms an embedded source bitstream into N packets of L symbols (for example, bytes) such that layers of decreasing importance are protected with increasingly weaker RS codes [54][57][62]. Borrowing notation from [62],a typical protection system [57][62] builds L segments $S_1, \ldots, S_L$, each of which consists of $k_i \in 1, \ldots, N$ $(i \in \{1, \ldots, L\})$ source symbols, and protects each segment

S_i with an (N, k_i) systematic RS code of maximal distance (see Fig. 25). For each $i \in \{1, \ldots, L\}$, let $f_i = N - k_i$ denote the number of RS parity symbols that protect segment S_i. An (N, k_i) systematic RS code can correctly recover the k_i information symbols as long as at least k_i symbols are not erasures. Therefore, if k packets of N are lost, then all the segments that contain at most $N - k$ source symbols can be recovered. Since the layers of the embedded source bitstream have decreasing importance, increasingly weaker RS codes should be used to provide UEP, that is, $f_1 \geq f_2 \geq \ldots \geq f_L$. Hence if at most f_i packets are lost, the receiver can decode at least the first i segments. Let $p_N(x = n)$ denote the probability of losing exactly n packets of N, the probability that the receiver can correctly recover segment S_i is the probability of losing at most f_i packets, i.e., $p_N(x \leq f_i) = \Sigma_{x=0}^{f_i} p_N(x)$.

Denote by $\mathcal{F}$ the set of L-tuples $\{f_1, f_2, \ldots, f_L\}$ where $0 \leq f_i \leq N - 1$ for $i = 1, \ldots, L$ and $f_1 \geq f_2 \geq \ldots \geq f_L$. The UEP scheme for the source packets is described by specifying a code allocation policy [55]. A code allocation policy F allocates an $(N, N - f_i)$ systematic RS code to the ith segment of the bitstream. The distortion optimal L-segment UEP scheme is a code allocation policy $F = (f_1, \ldots, f_L) \in \mathcal{F}$ that minimizes the expected distortion

$$E_N d(F) = \Sigma_{i=0}^{N} P_i(F) d_i(F), \tag{5.1}$$

where $P_0(F) = P(X > F_1)$, $P_i(F) = P(f_{i+1} < X \leq f_i)$ for $1 \leq i \leq L - 1$, $P_L(F) = P(X \leq f_L)$, $d_0(F) = d_0$ is the source variance, and for $i \geq 1$, $d_i(F)$ is the reconstruction distortion using the first i packets. Note that $P_i(F) = \Sigma_{x=f_{i+1}+1}^{f_i} p_N(x)$, which is the probability that exactly the first i packets are correctly received. The rate-distortion data should be computed beforehand by decoding the scalable bitstream at different rates.

B. Wyner-Ziv Video Coding Using RS Codes for Packet Erasure Channel

Treating a standard coded video as the base layer (or side information), Wyner-Ziv video coding generates a layered Wyner-Ziv bitstream of the original video sequence to enhance the base layer such that it is still decodable with commensurate qualities at rates corresponding to layer boundaries. It compresses a video sequence into two bit streams: one for the low-quality, non-scalable base layer generated by the classic motion-compensated DCT approach and another for the scalable enhancement layer, which is formed by applying the DCT, NSQ and LDPC code based bit plane coding for SWC (Fig. 26). The NSQ with nesting ratio N generates an output bitstream consists $M = \log_2(N)$ bit planes $B_0, \ldots, B_{M-1}$ and the SWC produces one syndrome layer for each bit plane B_i using LDPC code with rate corresponding to the conditional entropy $H(B_i|Y, B_{i-1}, \ldots, B_0)$. The bit plane coding produces an embedded bit stream that can be arbitrarily truncated at layer boundaries to fit the available channel bandwidth. For scalable video coding, it is reasonable to assume that the bandwidth is large enough that the base layer can be successfully received, hence we only focus on the transmission of the enhancement layer over packet erasure channel.

The transmission process proceeds as follows. The embedded enhancement bitstream is sent as N packets of L symbols, where each segment S_i consists of k_i source symbols and $f_i = N - k_i$ RS parity symbols to provide loss protection, $i \in \{1, \ldots, L\}$. The rate of the RS code for each segment is chosen according to some code assignment policy. The receiver tries to recover the source packets based on the correctly received packets from the packet erasure channel. For embedded source coders it is often reasonable to assume that if a source packet is decoded erroneously by the receiver, then the subsequent source packets cannot improve the quality of the source.

With our UEP structure, this assumption is naturally satisfied since the lower layers can not be successfully recovered if at least one of the higher layers can not. Hence, at any stage in transmission, the source is reconstructed only from the decoded bit stream up to the first source packet that contains a detectable error or irrecoverable erasure.

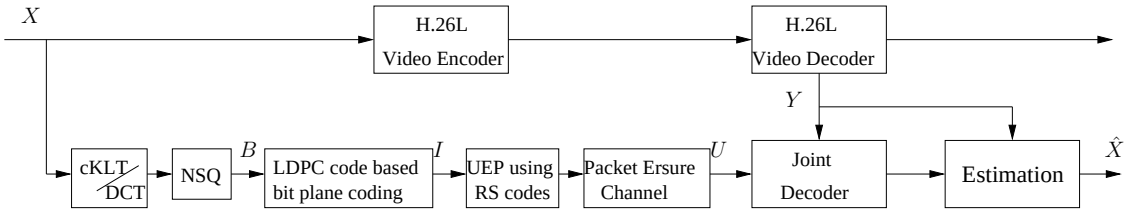


Fig. 26. Wyner-Ziv video coding using RS codes for packet erasure channel.

To design the distortion optimal L-segment UEP scheme, note that since the bits in one syndrome layer are equally important for decoding, they are protected with RS codes of equal strength. Therefore, the UEP scheme can be denoted by a rate allocation $F = \{f_1, f_2, \ldots, f_M\}$, where f_i is the parity bits used to protect each segment of the ith syndrome layer, $f_i \in \{0, \ldots, N-1\}$ for $i = 1, \ldots, M$ and $f_1 \geq f_2 \geq \ldots \geq f_M$. The probability of receiving the first i syndrome layers is then $P_i(F) = p_N(x \leq f_i) = \Sigma_{x=0}^{f_i} p_N(x)$. In addition, the distortion on receiving the first i syndrome layers is denoted by $d_i(F)$. Therefore, the distortion optimal L-segment UEP scheme is a code allocation policy $F = (f_1, \ldots, f_M)$ that minimizes the expected distortion

$$E_N d(F) = \Sigma_{i=0}^{N} P_i(F) d_i(F). \tag{5.2}$$

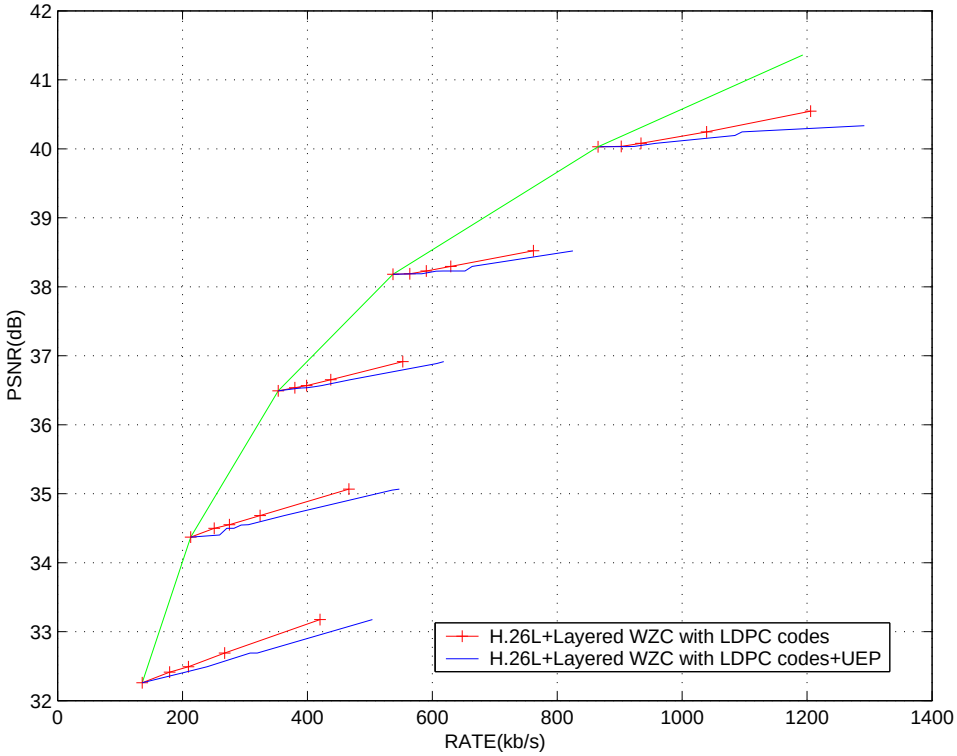


Fig. 27. Performance of Wyner-Ziv video coding with RS codes over packet erasure channel with $p = 0.20$ for "Foreman".

C. Experiment Results of Coding Efficiency

In our simulations, scalable bitstreams are generated off-line, with Wyner-Ziv video coding using LDPC coded based bit plane coding for SWC. Each source is encoded into an emdedded bitstream only once and the operational R-D data are stored as look-up tables. The transmission "server" computes the rate allocation based on the R-D data, the channel condition and the target transmission rate. Then it generate the RS codes and packetized the truncated bitstream into equal-length packets. We choose the packet length to be 48 bytes.

One hundred frames are compressed with a frame rate of 30Hz. For each of these sequences, the first frame is coded as I frame, and all the subsequent frames as P frames by H.26L. The encoder separate the video sequence into a series of group

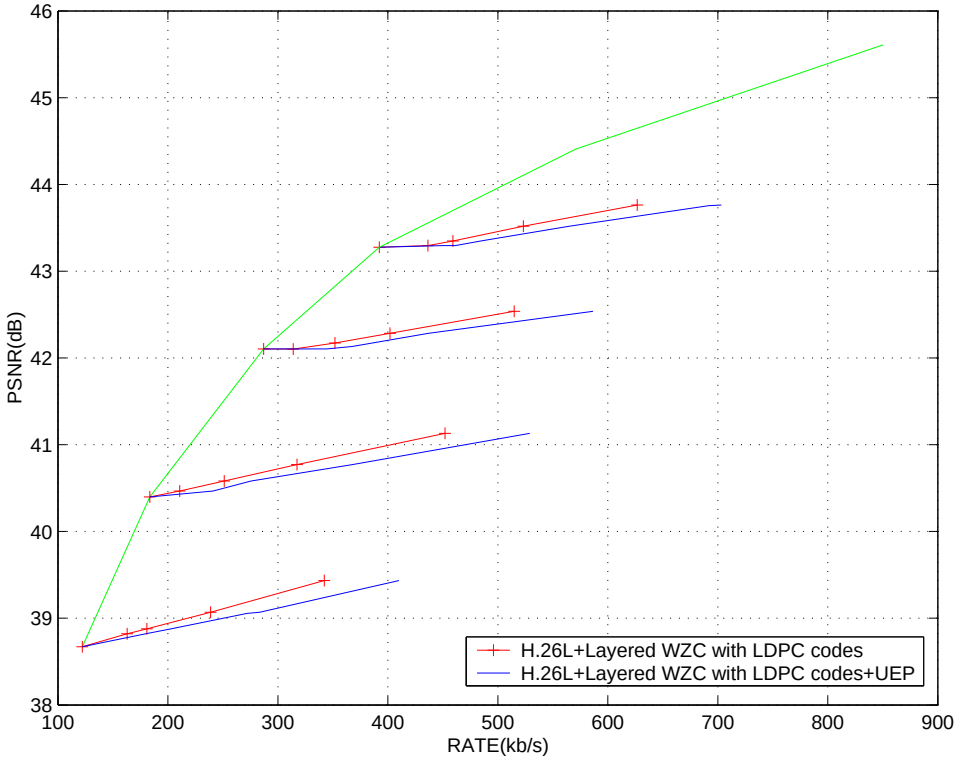


Fig. 28. Performance of Wyner-Ziv video coding with RS codes over packet erasure channel with $p = 0.20$ for "Mother_daughter".

of frames (GOF), each of which contains 20 frames. Each GOF is then encoded into an embedded bit stream. For simplicity, we use the average R-D data for all GOFs. We assume that the base layer bitstream is transmitted losslessly and we only consider the transmission of the enhancement layer bitstream. We design the UEP scheme for the enhancement layer bitstream to generate a fixed number of equal-length packets according to the target transmission rate. The packet mean loss rate of the erasure channel is 0.20. The goal of the optimal UEP scheme was to maximize the expected PSNR. Thus, instead of the cost function (5.2), we used the objective function $\Sigma_{i=0}^{L} P_i(F) PSNR_i(F)$, where $PSNR_i(F)$ is the PSNR corresponding to the first i syndrome layers. Fig. 27 represents the performance in terms of rate vs. average PSNR for "Foreman". To achieve the same PSNR performance as in noiseless channel

case, an extra 0.12 b/s is required to provide erasure protection. The performance for "Mother_daughter" sequence is plotted in Fig. 28. The extra bit rate required to achieve the same performance as in noiseless channel case is 0.16 b/s.

CHAPTER VI

CONCLUSION

We have introduced the first practical layered video coder based on the Wyner-Ziv coding principle and addressed its error robustness. Treating a standard coded video as the base layer (or side information), a layered Wyner-Ziv bitstream of the original video sequence is generated to enhance the base layer such that it is still decodable with commensurate qualities at rates corresponding to layer boundaries. The simple Wyner-Ziv video encoder consists of the DCT, NSQ and LDPC codes based bit plane coding for SWC. Our proposed system is similar to FGS coding in the sense of an embedded enhancement layer with good R-D performance. However, the main motivation of FGS coding is scalable coding for error robustness while assuming that the base layer (with heavy error protection) is always available loselessly at the decoder/receiver. With Wyner-Ziv coding, this stringent requirement can be loosened somewhat as an error-concealed version of the base layer can still be used in the joint Wyner-Ziv decoder. Experiment results show the superior error-robustness of Wyner-Ziv video coding with noisy side information at the decoder over FGS coding .

For Wyner-Ziv coding over discrete noisy channels, additional channel coding is needed to protect the Wyner-Ziv bitstream from channel errors. WZC can be thought of as a quantizer followed by SWC, which can be viewed as a channel coding problem. The channel coding component used for error protection can be combined with the SWC component in a joint design. We propose a JSCC framework for Wyner-Ziv video coding over BSC based on the idea of two equivalent channels [23]. Compared with the traditional video transmission systems which use feedback and retransmission of lost packets based on FEC, our system enables one integrated design of both the systematic part and the parity part of the IRA codes for different channel conditions

and provides good R-D performance.

For video streaming applications where the transmission channel is packet based, a Wyner-Ziv video coding scheme using UEP of embedded data for packet erasure channel is presented. UEP generates distortion optimal packing scheme that minimizes the expected distortion at a target transmission rate. With a combined design of WZC and UEP, we obtain an efficient video coding and transmission system at target transmission rate over packet erasure channels.

These initial results are encouraging, but much remains to be done. For example, as stated in chapter III, more accurate statistical distribution modeling is necessary to improve the code design for WZC. Capacity achieving channel code that performs well with short block length is desired for SWC. Finally, our current implementation of layered Wyner-Ziv coding only allows decoding at layer boundaries – decoding at the middle of a layer (bit plane) will suffer a huge performance loss as unavailable bits in the layer have to be treated as erasures. This is due to the limitation of the LDPC/IRA codes used for SWC. *Progressively decodable* channel (or Slepian-Wolf) codes such as Tornado codes [64] should be studied to allow fine granularity scalability of Wyner-Ziv video coding.

REFERENCES

[1] B. Girod, A. Aaron, S. Rane, and D. Rebollo-Monedero, "Distributed video coding," *Proc. of the IEEE*, Special Issue on Video Coding and Delivery, 2004 (to appear).

[2] A. Sehgal, A. Jagmohan, N. Ahuja, "Wyner-Ziv Coding of Video: Applications to Error Resilience," *IEEE Trans. Multimedia*, pp. 249-258, April 2004,.

[3] R. Puri and K. Ramchandran, "PRISM: A video coding architecture based on distributed compression principles," submitted to *IEEE Trans. Image Processing*, 2003.

[4] Y. Sterinberg and N. Merhav, "On successive refinement for the Wyner-Ziv problem," *IEEE Trans. Inform. Theory* (to appear).

[5] S. Cheng and Z. Xiong, "Successive refinement for the Wyner-Ziv problem and layered code design," in Proc. DCC'04, Snowbird, UT, April 2004, pp. 531.

[6] M. Gastpar, P. Dragotti, M. and Vetterli, "The distributed, partial, and conditional Karhunen-Loeve transforms," in Proc. DCC'03, Snowbird, UT, March 2003, pp. 283-292.

[7] T. Chu and Z. Xiong, "Coding of Gauss-Markov sources with side information at the decoder," in Proc. IEEE Workshop on Statistical Signal Processing, St. Louis, MO, September 2003, pp. 26-29.

[8] MPEG-4 Video VM, ver. 13.0, ISO/IEC JTC 1/SC29/WG11 N2687, March 1999.

[9] T. Wiegand, G. Sullivan, G. Bjintegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Trans. Circuits and Systems for Video Tech.,* vol. 13, pp. 560-576, July 2003.

[10] J.D. Slepian and J.K. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inform. Theory*, vol. 19, pp. 471-480, July 1973.

[11] A. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. Inform. Theory*, vol. 22, pp. 1-10, January 1976.

[12] R. Zamir, S. Shamai, and U. Erez, "Nested Linear/Lattice Codes for Structured Multiterminal Binning", *IEEE Transactions on Information Theory*, special A.D. Wyner issue, pp. 1250-1276, June 2002.

[13] S. Servetto, "Lattice quantization with side information, " in Proc. DCC'00, Snowbird, UT, March 2000, pp. 510-519.

[14] X. Wang, and M.T. Orchard, "Design of trellis codes for source codng with side information at the decoder," in Proc. DCC'01, Snowbird, UT, March 2001, pp. 361-370.

[15] J. Chou, S. Pradhan and K. Ramchandran, "Turbo and trellis-based constructions for source coding with side information," in Proc. DCC'03, Snowbird, UT, March 2003, pp. 33-42.

[16] S. Pradhan and K. Ramchandran, "Distributed source coding using syndromes (DISCUS): Design and construction," *IEEE Trans. Inform. Theory*, vol. 49, pp. 626-643, March 2003.

[17] Z. Xiong, A. Liveris, S. Cheng, and Z. Liu, "Nested quantization and Slepian-Wolf coding: A Wyner-Ziv coding paradigm for i.i.d. sources," in Proc. IEEE Workshop on Statistical Signal Processing, St. Louis, MO, September 2003, pp. 399-402.

[18] Y. Yang, S. Cheng, Z. Xiong, and W. Zhao, "Wyner-Ziv coding based on TCQ and LDPC codes," in Proc. of 37th Asilomar Conf. on Signals, Systems, and Computers, Pacific Grove, CA, November 2003, pp. 825-829.

[19] W. Li, "Overview of fine granularity scalability in MPEG-4 video standard", *IEEE Trans. Circuits and Systems for Video Tech.,* vol. 11, pp. 301-317, March, 2001.

[20] Y. He, R. Yan, F. Wu, and S. Li, "H.26L-based fine granularity scalable video coding", in IEEE Int. Symp. on Circuits and Systems (ISCAS), Phoenix, AR, May 2002, vol. 4, pp 548-551.

[21] W. Equitz and T. Cover, "Successive refinement of information," *IEEE Trans. Inform. Theory*, vol. 37, pp. 269-274, March 1991.

[22] H. Jin, A. Khandekar, and R. McEliece, "Irregular repeat-accumulate codes," in Proc. 2nd Int. Symp. on Turbo codes, Brest, France, Sept. 2000, pp. 1-8.

[23] A. Liveris, Z. Xiong, and C. Georghiades, "Joint source-channel coding of binary sources with side information at the decoder using IRA codes," in Proc. Multimedia Signal Processing Workshop, St. Thomas, US Virgin Islands, December 2002, pp. 53-56.

[24] R. Gallager, *Low Density Parity Check Codes*, Cambridge, MA: MIT Press, 1963.

[25] T.Cover and J.Thomas, *Elements of Information Theory*, New York: Wiley, 1991.

[26] A. Wyner, "Recent results in the Shannon theory," *IEEE Trans. Inform. Theory*, vol. 20, pp. 2-10, January 1974.

[27] C. Berrou and A. Glavieux, "Near optimum error correcting coding and decoding: turbo-codes," *IEEE Trans. Communications,* vol. 44, pp. 1261-1271, October 1996.

[28] J. Garcia-Frias and Y. Zhao, "Compression of correlated binary sources using turbo codes," *IEEE Communication Letters,* vol. 5, pp. 417-419, October 2001.

[29] J. Bajcsy and P. Mitran, "Coding for the Slepian-Wolf problem with turbo codes," in Proc. GlobeCom'01, San Antonio, TX, November 2001, pp. 1400-1404.

[30] A. Aaron and B. Girod, "Compression with side information using turbo codes," in Proc. DCC'02, Snowbird, UT, April 2002, pp. 0252-0261.

[31] A. Liveris, Z. Xiong and C. Georghiades, "Distributed compression of binary sources using conventional parallel and serial concatenated convolutional codes," in Proc. DCC'03, Snowbird, UT, March 2003, pp. 193-202.

[32] A. Liveris, Z. Xiong and C. Georghiades, "Compression of binary sources with side information at the decoder using LDPC codes," *IEEE Communications Letters*, vol. 6, pp. 440-442, October 2002.

[33] C. Lan, A. Liveris, K. Narayanan, Z. Xiong, and C. Georghiades, "Slepian-Wolf coding of multiple M-ary sources using LDPC codes," in Proc. DCC'04, Snowbird, UT, March 2004, pp. 549.

[34] A. Liveris, C. Lan, K. Narayanan, Z. Xiong, and C. Georghiades, "Slepian-Wolf coding of three binary sources using LDPC codes," in Proc. 3rd Intl. Symp. Turbo codes, Brest, France, September 2003, pp. 63-66.

[35] Zamir, R., "The rate loss in the Wyner-Ziv problem," *IEEE Trans. Inform. Theory*, vol. 42, pp. 2073-2084, November 1996.

[36] S.S. Pradhan, J. Chou, and K. Ramchandran, "Duality between source coding and channel coding and its extension to the side information case", *IEEE Trans. Inform. Theory*, vol. 49, pp. 1181-1203, May 2003.

[37] M. Marcellin and T. Fischer, "Trellis coded quantization of memoryless and Gaussian-Markov sources", *IEEE Trans. Communications,* vol. 38, pp. 82-93, January, 1990.

[38] J. Conway and N. Sloane, *Sphere Pacings, Lattices and Groups*, Springer Verlag, New York, 1998.

[39] R. Zamir, S. Shamai, and U. Erez, "Nested linear/lattice codes for structured multiterminal binning," *IEEE Trans. Inform. Theory*, vol. 48, pp. 1250-1276, June 2002.

[40] 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. DCC'04, Snowbird, UT, March 2004, pp. 322-331.

[41] Lori A. Dalton, "Analysis of 1-D nested lattice quantization and Slepian-Wolf coding for Wyner-Ziv coding of i.i.d. sources", Project report for ELEN 663, Texas A&M University, May 2003 (personal collection).

[42] J. M. Shapiro, "Embedded image coding using zerotrees of wavelet coefficients," *IEEE Trans. Signal Processing,* vol. 41, pp. 3445-3463, December 1993.

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

[44] B.-J. Kim, Z. Xiong, and W. Pearlman, "Very low bit-rate embedded video coding with 3-D set partitioning in hierarchical trees (3-D SPIHT)," *IEEE Trans. Circuits and Systems for Video Tech.,* vol. 20, pp. 1365-1374, December 2000.

[45] D. MacKay, "Good error-correcting codes based on very sparse matrices," *IEEE Trans. Inform. Theory*, vol. 45, pp. 399-431, March 1999.

[46] D. Reininger and J. Gibson, "Distribution of the two-dimensional DCT coefficients," *IEEE Trans. Communications*, vol. 31, pp. 835-839, June 1983.

[47] T. Richardson, M. Shokrollahi, and R. Urbanke, "Design of capacity-approaching irregular low-density parity-check codes," *IEEE Trans. Inform. Theory*, vol. 47, pp. 619–637, February 2001.

[48] S. Chung, T. Richardson, and R. Urbanke, "Analysis of sum-product decoding of low-density parity-check codes using a Gaussian approximation," *IEEE Trans. Inform. Theory*, vol. 47, pp. 657-670, February 2001.

[49] S. Shamai, S. Verdu, and R. Zamir, "Systematic lossy source/channel coding," *IEEE Trans. Inform. Theory*, vol. 44, pp. 564-579, March 1998.

[50] S. Shamai (Shitz) and S. Verd, "Capacity of channels with side information," *European Transactions on Telecommunications*, vol. 6, no. 5, pp. 587-600, Sept.-Oct. 1995.

[51] S. Shamai and S. Verdu, "Capacity of channels with uncoded-message side-information," in Proc. Intl. Symp. on Inform. Theory, Whistler, Canada, September 17-22, 1995, pp. 7.

[52] S.-Y. Chung, *On the Construction of Some Capacity-Approaching Coding Schemes,* Ph.D. dissertation, Massachusetts Institute of Technology, 2000.

[53] J. Hagenauer, "Rate-compatible Punctured Convolutional Codes (RCPC Codes) and their applications," *IEEE Trans. Communications*, vol. 36, pp. 389-400, April 1988.

[54] R. Puri, K. Ramchandran, 'Multiple description coding using forward error correction codes," in Proc. 33rd Asilomar Conf. on Signals, Systems, and Computers, Piscataway, NJ, October 1999, vol. 1, pp. 342-346.

[55] V. Chande and N. Farvardin, "Progressive transmission of images over memoryless noisy channels," *IEEE Journal on Selected Areas in Communications*, vol. 16, pp. 850-860, June 2000.

[56] A. Mohr, E. Riskin, and R. Ladner, "Unequal loss protection: graceful degradation of image quality over packet erasure channels through forward error correction," *IEEE Journal on Selected Areas in Communications*, vol. 18, pp. 819-828, June 2000.

[57] A. Mohr, R. Ladner, E. Riskin, "Approximately optimal assignment for unequal loss protection," in Proc. IEEE ICIP-2000, Vancouver, Canada, Sept. 2000, vol. 1, pp. 367-370.

[58] P. A. Chou, A. E. Mohr, A. Wang, and S. Mehrotra, "Error control for receiver-driven layered multicast of audio and video," *IEEE Trans. on Multimedia*, vol.

3, pp. 108-122, March 2001.

[59] B. Banister, B. Belzer and T. Fischer, "Robust image transmission using JPEG2000 and turbo-codes," *IEEE Signal Processing Letters*, vol. 9, pp. 117-119, April 2002.

[60] A. Nosratinia, J. Lu, and B. Aazhang, "Source-channel rate allocation for progressive transmission of images," *IEEE Trans. Communications*, vol. 51, pp. 186-196, February 2003.

[61] V. Stankovic, R. Hamzaoui, Y. Charfi, and Z. Xiong, "Real-time unequal error protection algorithms for progressive image transmission," *IEEE JSAC on Recent Advances in Wireless Multimedia*, vol. 21, pp. 1526-1535, December 2003.

[62] V. Stankovic, R. Hamzaoui, and Z. Xiong, "Real-time error protection algorithms of embedded codes for packet erasure and fading channels," *IEEE Trans. Circuits and Systems for Video Tech.*, 2004 (to appear).

[63] C. Lan, T. Chu, K. Narayanan, and Z. Xiong, "Scalable image and video transmission using irregular repeat accumulate (IRA) codes with fast algorithm for optimal unequal error protection," *IEEE Trans. Communications*, vol. 52, pp. 1092-1101, July 2004.

[64] M. Luby, M. Mitzenmacher, and A. Shokrollahi, "Analysis of random processes via and-or tree evaluation", in Proc. of 9th Annual ACM-SIAM Symposium on Discrete Algorithms, San Francisco, California, January 25-27, 1998, pp. 364-373.

APPENDIX A

THE DETAILS OF IMPLEMENTATION IN CHAPTER III.

Table II. The theoretical rate limit and the actual LDPC code rate used for each bit plane after the NSQ of the DC component (a) and the first two AC components (b, c) of the DCT coefficients.

ith bit plane	$H(B_i\|Y, B_{i-1}, \ldots, B_0)$	LDPC code rate
0	0.03433	0.94
1	0.03884	0.94
2	0.10157	0.85
3	0.19593	0.73

(a)

ith bit plane	$H(B_i\|Y, B_{i-1}, \ldots, B_0)$	LDPC code rate
0	0.00178	0.99
1	0.03768	0.94
2	0.02532	0.96
3	0.16754	0.77

(b)

ith bit plane	$H(B_i\|Y, B_{i-1}, \ldots, B_0)$	LDPC code rate
0	0.02785	0.95
1	0.00488	0.98
2	0.03240	0.94
3	0.44291	0.49

(c)

Table III. LDPC code profiles. $\lambda(x) = \Sigma_{i=2}^{J} \lambda_i x^{i-1}$, $\rho(x) = 0.5x^{\alpha-1} + 0.5x^{\alpha}$.

rate	0.50	0.51	0.52	0.53	0.54	0.55	0.56	0.57	0.58	0.59
λ_2	0.2086	0.2416	0.2163	0.2199	0.2002	0.2282	0.2076	0.2392	0.2392	0.2444
λ_3	0.1448	0.1800	0.1555	0.1574	0.1495	0.1800	0.1601	0.1915	0.1915	0.1925
λ_4			0.0177	0.0388						
λ_5	0.1302		0.0683	0.0162	0.0947		0.1048	0.0201	0.0201	0.0782
λ_6		0.2406			0.0534	0.2074		0.2271	0.2271	0.1554
λ_7		0.0123	0.0942	0.2350		0.0518				
λ_8			0.1154				0.1958			
λ_9	0.1495				0.0788					
λ_{10}	0.0454				0.0974					
λ_{14}										0.2336
λ_{15}								0.0824	0.0824	0.0960
λ_{16}								0.2397	0.2397	
λ_{17}		0.0947								
λ_{18}		0.2308				0.1230				
λ_{19}						0.2095				
λ_{21}				0.0249						
λ_{22}				0.3078						
λ_{23}			0.2110				0.0542			
λ_{24}			0.1216				0.2775			
λ_{27}					0.0093					
λ_{28}	0.3215				0.3168					
α	9	8	9	9	10	9	10	9	9	9

(a)

rate	0.60	0.61	0.62	0.63	0.64	0.65	0.66	0.67	0.68	0.69
λ_2	0.2576	0.1901	0.2394	0.2205	0.2256	0.2103	0.2154	0.2029	0.1927	0.1845
λ_3	0.1916	0.1567	0.1940	0.1937	0.1957	0.1915	0.1958	0.1901	0.1830	0.1723
λ_4	0.0423									
λ_5	0.1388	0.0882	0.1381		0.0443					0.0013
λ_6		0.0480	0.0754	0.2374	0.1946	0.1838	0.2470	0.1627	0.1516	0.2284
λ_7				0.0139		0.0790		0.1027	0.1070	
λ_9		0.1784								
λ_{10}		0.0002								
λ_{11}	0.2987									
λ_{12}	0.0709									
λ_{13}			0.2832							
λ_{14}			0.0699							
λ_{15}					0.3096					
λ_{16}				0.1378	0.0302		0.2345			
λ_{17}				0.1967			0.1074			
λ_{18}						0.3354		0.0839		
λ_{19}								0.2577		
λ_{20}									0.3656	0.4135
λ_{26}		0.1578								
λ_{27}		0.1804								
α	9	12	10	11	11	12	12	13	14	15

(b)

rate	0.70	0.71	0.72	0.73	0.74	0.75	0.76	0.77	0.78	0.79
λ_2	0.1892	0.2116	0.2190	0.2103	0.2382	0.2296	0.2377	0.2313	0.2266	0.2348
λ_3	0.1855	0.2054	0.2080	0.2062	0.2101	0.2101	0.2272	0.2201	0.2117	0.2519
λ_4					0.1213	0.0873	0.1006	0.1092	0.1235	0.0657
λ_5		0.0192	0.1089	0.0615		0.0556				
λ_6	0.1213	0.2213	0.1047	0.1667						
λ_7	0.1474									
λ_8							0.0396			0.2656
λ_9					0.2838		0.3949	0.3477	0.2199	0.1820
λ_{10}					0.1466	0.4173		0.0917	0.2183	
λ_{13}			0.3593							
λ_{14}				0.3002						
λ_{15}		0.3425		0.0551						
λ_{20}	0.1692									
λ_{21}	0.1874									
α	15	14	14	15	14	15	15	16	17	17

(c)

rate	0.80	0.81	0.82	0.83	0.84	0.85	0.86	0.87	0.88	0.89
λ_2	0.2056	0.2291	0.1852	0.1866	0.2196	0.2137	0.2100	0.2006	0.2275	0.2462
λ_3	0.2211	0.2556	0.2141	0.2204	0.2562	0.2482	0.2431	0.2480	0.3173	0.3185
λ_4		0.0626			0.0673	0.0795	0.0906	0.0201		
λ_5	0.1342							0.1099		
λ_6	0.0549		0.2374	0.2352						0.3586
λ_7									0.3258	0.0767
λ_8		0.2620			0.1967	0.0695			0.1294	
λ_9		0.1906			0.2601	0.3889	0.4485			
λ_{10}							0.0078	0.4214		
λ_{14}				0.1000						
λ_{15}			0.2373	0.2577						
λ_{16}			0.1260							
λ_{20}	0.3843									
α	20	19	24	25	23	25	27	30	29	30

(d)

VITA

Qian Xu was born in Jingzhou, China in 1982. She received the B.S. degree in Special Class for the Gifted Young from University of Science&Technology of China in 2002 and M.S. degrees in electrical engineering from Texas A&M University in 2004. She is currently a Ph.D. student under Dr. Zixiang Xiong in the Wireless Communications Laboratory at Texas A&M University.

Upon admittance to Texas A&M, Ms. Xu was a recipient of the TxTEC Fellowship in September 2002. From 2001 to 2002, she worked as a student intern in the multimedia group at Microsoft Research Asia, Beijing, China.

Her areas of interest include distributed video and stereo coding, video compression and digital communication.

Ms. Xu can be reached by the following address:

Jiangling Middle School,
Jingzhou City, Hubei Province,
P.R.China, 434100

The typist for this thesis was Qian Xu.