5-1-2009

ROBUST IMAGE AND VIDEO CODING WITH ADAPTIVE RATE CONTROL

Hongqiang Wang
University of Nebraska at Lincoln, why_whq@hotmail.com

ROBUST IMAGE AND VIDEO CODING WITH ADAPTIVE RATE CONTROL

By

Hongqiang Wang

A DISSERTATION

Presented to the Faculty of

The Graduate College at the University of Nebraska

In Partial Fulfillment of Requirements

For the Degree of Doctor of Philosophy

Major: Interdepartmental Area of Engineering

Under the Supervision of Professor Khalid Sayood

Lincoln, Nebraska

May, 2009

ROBUST IMAGE AND VIDEO CODING WITH ADAPTIVE RATE CONTROL

Hongqiang Wang, PhD
University of Nebraska, 2009

Advisor: Khalid Sayood

This dissertation is focused on the problem of rate allocation in a resource constrained environment and robustness of video coding over noise channels. A new rate allocation scheme that combines the traditional PCRD-Opt algorithm with the ρ domain analysis is proposed for wavelet based image coders. The proposed scheme provides competitive performance as compared with the optimal PCRD-Opt algorithm, while with significant reduction on complexity and computational costs.

Rate allocation is developed for the **Region-Of-Interest (ROI)** in the wavelet transform domain for image coding. A recursive region growing method that determines the ROI in the transform domain is proposed. With excellent coding performance, robustness of the coder is improved as well because the ROI is allocated more bit resources and better protected. In addition, as intra-frame video coding is less vulnerable to errors than inter-frame video coding, we proposed rate allocation methods for several intra-frame video coding scenarios based on the proposed rate allocation algorithm for image coding.

A coding method based on the principles of distributed source coding in the wavelet transform domain is proposed in order to achieve a good compromise between compression performance and robustness. 8×8 blocks in the wavelet transform domain are classified and coded by either distributed source coding or intra-frame video coding depending on the amount of correlation between co-located blocks in two consecutive frames. The regions that contain motion blocks are extracted and then intra coded

with higher priority and more rate resources. The background regions are coded based on the principle of distributed video coding. The approach exploits inter-frame redundancy without explicitly using inter-frame video coding and demonstrated both a better robustness than the traditional intra-frame video coder H.263+ and improved compression performance over the image coder based intra-frame video coding.

To Hui Zheng, Xindong, and Sophia

Acknowledgements

I would like to thank Prof. Khalid Sayood, my supervisor, for his inspiring guidance and consistent support throughout my research. I am especially thankful to him for his guidance through the early months of confusion and painstaking effort in correcting my thesis. The research assistantship sponsored by the NASA was crucial to the successful completion of this work.

Prof. Michael Hoffman has encouraged me to think deeply and helped me improve the writing and organization of my thesis. Thanks also go to my PhD committee members Prof. Sina Balker at the department of electrical engineering, and Prof. Ashok Samal at the department of computer science for their help and support.

Dr. Pen-Shu Yeh at NASA has consistently supported my work and helped me achieve a better perspective of my own results. I also want to thank Dr. Rodney Grubbs from NASA's Marshall Space Flight Center for the HD video sequence for testing, and thank Dr. Aaron Kiley at NASA for fixing a few bugs in my programs.

I am grateful to my families for their *patience* and *love*. Without them this work would never have come into existence. I especially thank my wife, Hui Zheng, for her being with me at the hardest time of my life.

Finally, I want to thank David Russell, Mark Bauer, Ufuk, Baron, Ying Li, Dongshen Bi, Yongkui Wen, Jing Xu, Xinwang Zhang, Qin Chen, Youlu Wang, and Lin Zhu for their help and support.

Contents

List of Figures

LIST OF FIGURES

LIST OF FIGURES

List of Tables

Chapter 1

Introduction

The goal of data compression is to represent the information contained in data using fewer bits. The information can be recovered intact, in which case the compression is called lossless coding, or is partially lost while maintaining a certain level of fidelity in the reconstruction, in which case the compression is called lossy coding. Here the data includes digitized text, speech, audio, and video signal. Because of the explosive growth in the amount of multimedia data for transmission and storage, data compression has become very important in order to save bandwidth and increase storage capacity. Research and development on algorithms for data compression have achieved many great results and dramatically changed our lives in many aspects. Here are several examples:

- The cell phone we use employs low bit rate speech coders such as

CELP (2; 3; 4) and can achieve toll quality that once could only be achieved by high bit rate speech coders such as adaptive differential pulse-code-modulation (5; 6; 7; 8).

- We store and exchange photographs which are encoded using sophisticated image coding standards created by the *Joint Photographic Experts Group* (JPEG) (9; 10; 11; 12).

- We watch movies in digital format or participate in video conferencing using video coding standards such as H.261 (13) for low bit rate coding, H.263 (14) for video conferencing, MPEG-1 (15), MPEG-2 (16), MPEG-4 (17), and the most advanced video coding standard ITU-T H.264/Advanced Video Coding(AVC) (part 10 of MPEG-4) (18). The next generation video standard called H.265 is being developed (19; 20).

Data compression has reshaped the concept of visual communication and home entertainment, and greatly reduced the traffic load for internet transmission. It has been one of the leading factors in the revolution of information technology, and is continuing to play an important role in the future.

1.1 Objectives

The conventional rate control scheme used in wavelet based image coder JPEG2000 employs a post-completion rate allocation algorithm (the PCRD-Opt algorithm) to allocate bit resources and optimize the rate distortion performance (21; 22). Because the encoder has to complete the entire coding process, this method is not efficient and may not be the best choice for applications where there are restrictions on computational and memory resources.

Despite the fact that conventional motion-compensated inter-frame video coders are overwhelmingly used in practical coding applications, robustness is a major concern as inter-frame motion-compensated video coding is vulnerable to the error propagation if the bit stream is erroneously transmitted under error-prone channel conditions. Instead, many high-speed/high-definition video coding systems employ intra-frame video coding schemes that use existing image coders as basis for video compression, such as Motion JPEG/JPEG2000. Though its PSNR performance is generally inferior to that of inter-frame video coding, intra-frame video coding based on image coders has been widely accepted due to its low complexity, low computational costs, and robustness. However, intra-frame coding suffers from low compression performance.

The work presented in this dissertation attempt to solve the rate allocation problem in a resource constrained environment and enhance the robustness of the video coding based on existing wavelet image coders. We develop rate allocation schemes with limited and affordable complexity for wavelet based image coder, extend it to intra-frame based video coding, and then enhance its robustness using the correlation between neighboring frames. Several issues are addressed individually and then combined to achieve the final goal. Throughout this dissertation, an existing wavelet based image coder called CCSDS Recommendation for Space Data System Standards (23) is employed to exemplify our approaches. For simplicity, we call it the *Bit Plane Coder* (BPE). Note that the employment of this coder does not limit the generality of the methodologies for other transform based coders.

First, we focus on the rate allocation problem. We propose a new rate allocation method which combines the PCRD-Opt algorithm and the recently emerging ρ domain analysis (24; 25; 26; 27). This method reduces the complexity of the PCRD-Opt algorithm significantly by avoiding complete encoding and achieves excellent rate allocation performance.

Secondly, we propose an image classification scheme and define corresponding syntax elements to accommodate this coding scheme. This

part is fundamentally important as region classification is required in the robust video coding scheme. More specifically, we propose a region growing method in the wavelet transform domain to classify images into *Region-of-Interest* and *Region-of-non-Interest* based on the activity level of the blocks contained in the regions. Different from traditional region growing applications in pixel domain, the objective of this method is to improve rate-distortion performance and facilitate rate allocation. Therefore, the criteria of grouping coding blocks are mainly based on the factors that have influence on compression performance. The regions of interest are allocated more bit resources based on the proposed rate allocation scheme. A corresponding syntax element for transmission of a significant block map is defined to indicate the region each block belongs to. We show that competitive compression performance is achieved using the proposed scheme. In addition, the regions of interest may be transmitted with higher priorities and robustness can be enhanced.

Thirdly, we extend the image coder for video coding and propose rate allocation schemes to optimize the rate-distortion performance of the image coder based video coding. We focus on issues related to group of pictures (GOP) level rate allocation, frame level rate allocation, and two-level rate allocation.

Finally, after extending the wavelet based image coder to video coding, we propose a wavelet based robust video coding scheme using the principle of distributed video coding. The distributed video coding is an extended application of distributed source coding. The concept of distributed source coding is abstracted from the Slepian-Wolf theory (28) and Wyner-Ziv theory (29), which show that separate encoding of two correlated sources does not degrade the coding efficiency provided that the decoding is performed jointly. The conventional motion compensated video coders usually have a very high complexity at the encoder side due to motion estimation and motion compensation. Therefore, their hardware implementation for space applications is costly and infeasible. In this scheme, we use the concept of distributed source coding to develop an affordable and low complexity robust video coding scheme. The coding blocks in the wavelet transform domain are classified based on their correlation with co-located blocks in neighboring frames and then coded using the bit plane coding combined with the distributed source coding. This coding scheme demonstrates good robustness at moderate and high bit error rates and its complexity is much lower than conventional motion compensated video coders.

1.2 Outline

Chapter 2 provides a review of related research. The review aims to present basic knowledge on existing image and video coders, and research advances on rate control. The topics include rate-distortion optimization, video coding, distributed source coding, and a brief review of the BPE coder.

In Chapter 3 a new rate control scheme which combines the PCRD-Opt algorithm and ρ domain analysis is proposed. The PCRD-Opt algorithm was applied to the BPE coder and ρ domain analysis was adapted for this image coder. The performance of the new scheme is compared with that of the BPE coder using the PCRD-Opt algorithm and the constant bit rate BPE coder.

In Chapter 4 the BPE coder is extended to accommodate the concept of *Region-Of-Interest* (ROI) in the transform domain with adaptive rate control. A region growing method in the wavelet transform domain is proposed to find the ROI. Rate control methods to accommodate the ROI and the non-ROI are proposed and then applied to the BPE coder. Syntax modification of the BPE coder is suggested to incorporate the concept of ROI. We show comparable compression performance as compared with the regular BPE coder, with the advantage of easy

manipulation of the ROI with higher priority and better robustness.

Rate control methods are developed for frame-level and segment-level rate allocation for intra-frame video coding in Chapter 5. Unlike the traditional inter-frame video coding using motion-compensated predictive coding, intra-frame video coding requires much lower expense of time, complexity, and computational resources. In this chapter, the BPE coder is extended to intra-frame based video coding. The results show that the compression performance of BPE using the rate control algorithms is improved dramatically as compared with the regular constant bit rate BPE coder.

In Chapter 6 a wavelet image coding scheme using the concept of distributed source coding is presented. The purpose of the use of distributed source coding is to take advantage of temporal correlation in video sequences without using explicit inter-frame video coding. Blocks are classified based on their correlation with co-located blocks in past frames and then coded based on their classification. The compression performance shows improvement over intra-frame video coding. The complexity is significantly lowered as motion estimation and motion compensation are not required, and the dependency between frames in this distributed video coding is reduced as compared with the traditional

motion-compensated video coding, therefore, robustness is improved.

In Chapter 7 the work presented in this thesis is summarized. Recommendations are presented for further research work on relevant topics.

1.3 Main Contributions

The main contribution of the dissertation can be summarized as follows:

- Proposed a low complexity wavelet based robust video coding scheme based on the classification of coding blocks and distributed source coding.

- Proposed a new rate control scheme for the wavelet based image coder and significantly reduced the complexity of the conventional PCRD-Opt algorithm;

- Proposed a region growing method to segment images in the wavelet transform domain, and applied the proposed rate control method to optimize its compression performance;

- Extended the wavelet based BPE coder to intra-frame based video coding and proposed adaptive rate control algorithms to optimize its compression performance;

Source code and executable code of the BPE coder are available for download at the website http://hyperspectral.unl.edu.

Chapter 2

Literature Review

In this chapter several topics related to the research presented in this dissertation are reviewed. First basic information on image and video coding is presented in Section 2.1. A review of issues in rate control is presented in Section 2.2. An emerging video coding technique called distributed video coding is reviewed in Section 2.3, and in Section 2.4 we briefly discuss methods to improve robustness of video coding. After a review of the BPE coder in Section 2.5, conclusions are presented in Section 2.6.

2.1 Image and Video Coding

Generally, a lot of redundancy exists in natural images and video sequences. Data compression attempts to remove the redundancy and represent the original data using fewer bits. Depending on whether or

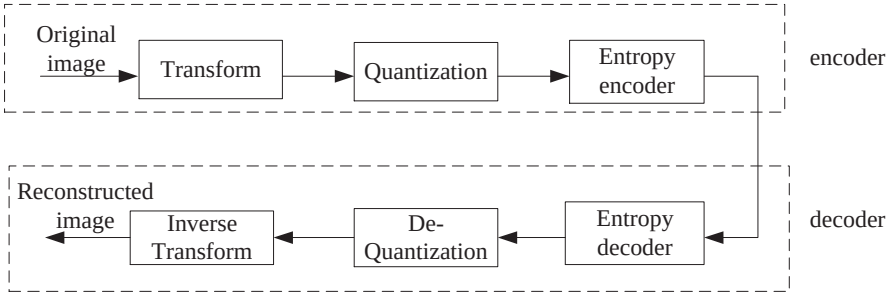


Figure 2.1: A generic transform based image coder

not there is information loss between the original and the reconstructed images and video sequences, compression can be classified as lossy or lossless. Most lossy compression methods are based on transform coding, such as the discrete cosine transform (DCT) in JPEG and the wavelet transform in JPEG2000. A generic transform coding scheme is shown in Figure 2.1, where the original signal is transformed first to remove redundancy. The resulting transform coefficients are then quantized and the quantization indices are entropy coded. The decoder takes steps in reverse order to reconstruct the image.

2.1.1 Wavelet Based Image Coding

Block-based *Discrete Cosine Transform* (DCT) has been used by many traditional compression schemes. Despite the advantages of compression schemes based on DCT such as simplicity and easy hardware implementation, at low bit rates their performance usually degrades rapidly and

noticeable and annoying "blocking artifacts" emerge. This is mainly because the input image is split into 8×8 disjoint blocks and each block is coded independently. The block-based DCT has an inherent discontinuity property across the block boundaries, which results in the "blocking artifacts" (30).

The Discrete wavelet transform (DWT) is a powerful tool in signal analysis and it overcomes some disadvantages of the Fourier transform. We know that Fourier transforms provide only frequency resolution but no time resolution. However, using a DWT we are able to represent a signal in the time and frequency domain at the same time. This property is especially useful for analysis and compression of non-stationary signal, such as natural images (31; 32).

In the past two decades, many powerful and sophisticated wavelet based image compression schemes have been proposed and standardized, such as the *Embedded Zero-tree Wavelet coder* (EZW) (33), the *Set Partitioning in Hierarchical Trees* (SPIHT) coder (34), EBCOT (35), the JPEG2000 coder (11), and the CCSDS Bit Plane Encoder (BPE) (23). The wavelet based image coders have several advantages over the traditional DCT based coders:

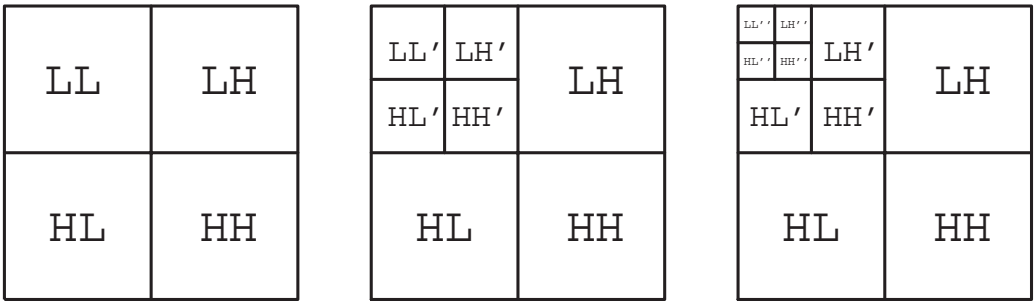
- The wavelet based image coders usually outperform the DCT based

image coders. For example, SPIHT is reported to have around a 3dB PSNR gain over JPEG on average (36).

- As the input image is not blocked, wavelet based compression schemes eliminate blocking artifacts at low bit rates and provide substantial improvement in picture quality.

- The wavelet based image coders can generate an embedded bit stream which facilitates progressive transmission of images. They are more robust under transmission and decoding errors.

- Due to their inherent multi-resolution nature, the wavelet based image coders are suitable for applications where high scalability and degradation tolerance are needed.



(a) 1-level decomposition (b) 2-level decomposition (c) 3-level decomposition

Figure 2.2: 3-level DWT decomposition

The wavelet transform is essentially a subband decomposition process and an image can be decomposed using low and high pass wavelet

filters in horizontal and vertical directions in multiple levels (37). There are several ways to do the decomposition and Figure 2.2 illustrates a widely used 3-level decomposition scheme in image compression. In this scheme, the first level decomposition generates four subbands, namely, low-low (LL), low-high (LH), high-low (LH), and high-high (HH) bands, as shown in Figure 2.2(a). The LL band is obtained by applying a low-pass filter to the rows and columns; the LH band is obtained by applying a low-pass filter to the rows and a high-pass filter to the columns; the HL band is obtained by applying a low-pass filter to the columns and a high-pass filter to the rows; the HH band is obtained by applying a high-pass filter to the rows and the columns. In the 2-level decomposition shown in Figure 2.2(b), the LL band from the first level is decomposed and replaced with four new subbands, while the other bands are left without further decomposition. The new subband is half the width and half the height of the LL subband from the previous level. Continuing to apply this decomposition to the LL band, a pyramidal multi-resolution decomposition structure is created. This structure is often referred to as the zero-tree structure, and we see that n levels of DWT decomposition can generate $3n + 1$ subbands. Figure 2.2(c) shows a 3-level decomposition, where ten bands are created.

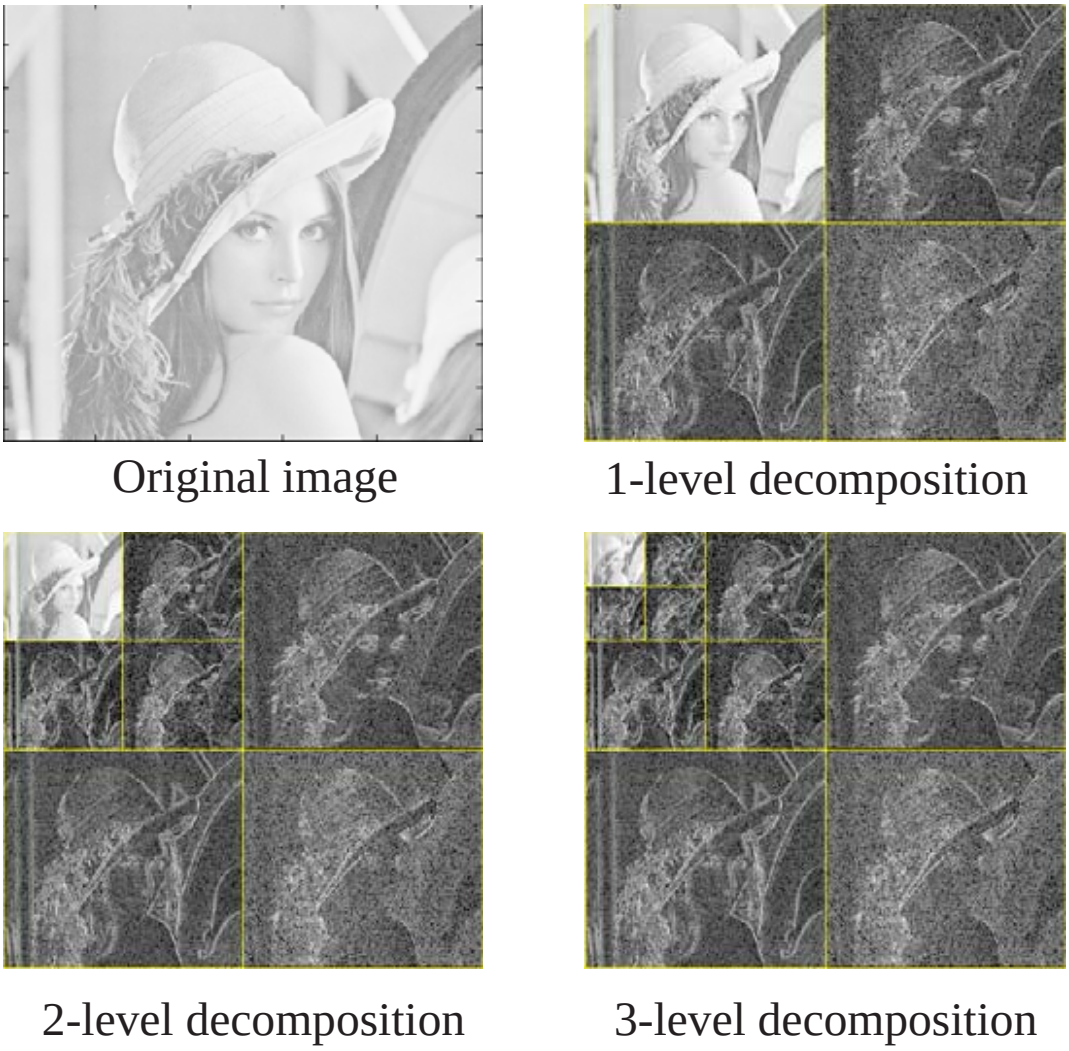


Figure 2.3: 3-level decomposition of Lenna image

The wavelet transform has excellent energy compaction and multi-resolution properties. Figure 2.3 shows the 3-level decomposition of Lenna image using Daubechies 4-tap filters (38). As we can see, most of the energy after wavelet decomposition is compacted in lower bands, i.e., the coefficients closer to the root of the structure have higher magnitudes than the rest of the coefficients. Therefore, the importance of those bands decreases from the top LL band to the HH band at the bottom. And we can observe self-similarity existing in the multi-resolution structure, and this implies that the bands are statistically correlated. The wavelet based image coders referred above take advantage of these properties and use intelligent methods to scan the quantization indices of the coefficients in a progressive manner. The resulting codewords can be effectively entropy coded. In Section 2.5 we will describe the wavelet based bit plane coder in detail and illustrate how the wavelet based image coder works.

2.1.2 Video Coding

For video sequences, there exist temporal, spatial, and spectral correlations. Temporal correlation exists between consecutive frames, while spatial correlation is between neighboring pixels in one frame. To achieve

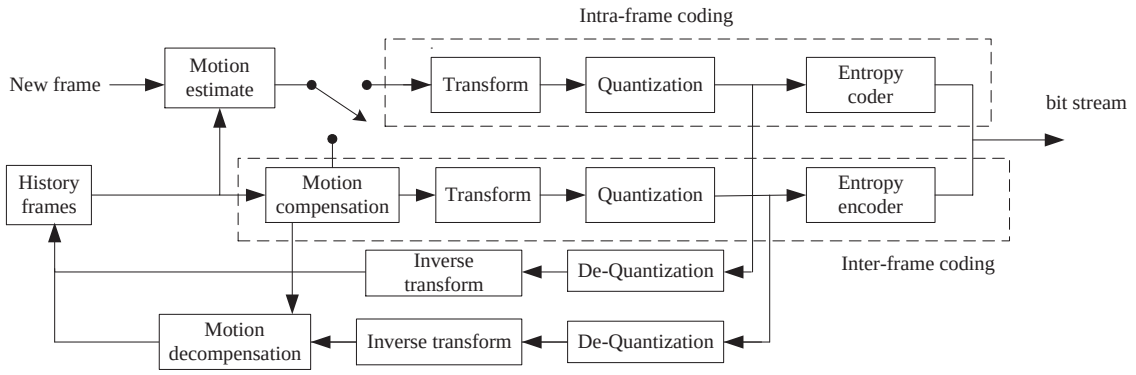


Figure 2.4: A generic transform based video encoder

good compression performance, video coders generally employ motion compensation based on inter-frame prediction to remove temporal correlation, and transform and predictive coding to remove spatial correlation.

Each frame is either intra-frame or inter-frame coded, as shown in a generic video coding block diagram in Figure 2.4. If a frame is coded without referring to other frames, it is called intra-frame coding, and the frame is called a key frame, or an I-frame. I-frames can be used as reference frames by other predictive frames, and help synchronize and prevent error propagation. A frame that is coded by referring to past frames for motion compensation and prediction is called predictive frame, or P-frames. As shown in Figure 2.5, the prediction residual after motion estimation and compensation between two similar blocks can be efficiently coded, along with motion vectors. Many video coders follow

a frame order in which multiple P frames are transmitted before an I frame: $IPPPPIPPPPI...P$. In this scenario, errors which occur in the bitstream for reference frames may propagate to the frames that use the erroneous frames as reference for motion compensation and prediction. This propagation may not be stopped until an I-frame is encountered to help regain the synchronization. Known as the "drifting effect", this type of error may severely degrade the reconstructed video sequences.

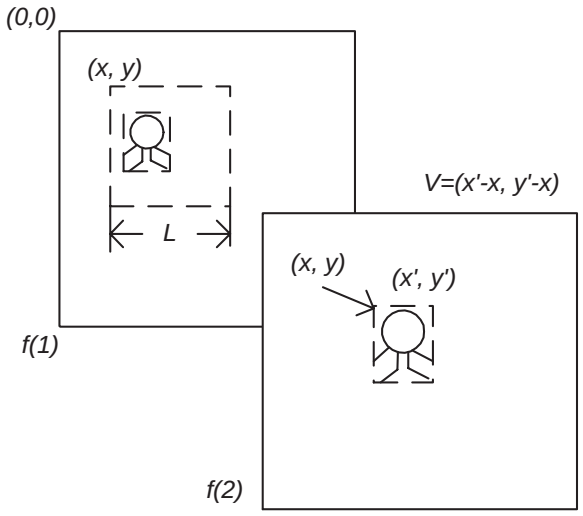


Figure 2.5: Motion estimate

Wavelet based video coding has been investigated extensively recently because it is relatively easier to achieve spatial, temporal, and SNR scalability and precise rate control using DWT based coders than it is using DCT based video coders (39; 40; 41; 42; 43; 44; 45; 46). In addition, wavelet video codes eliminate blocking artifacts, which are a well-known

issue in DCT based video coders.

Most wavelet video codes extend the existing DWT based image coders, such as SPIHT and EZW, to 3-D video coding with modifications to accommodate motion compensation. Kim. et al. proposed a 3-D SPIHT based low bit rate scalable video coding (39). Martucci et al. (40) proposed a low bit rate zero-tree based wavelet video coder, where an overlapping block motion compensation in combination with a discrete wavelet transform is followed by adaptive quantization and zero-tree entropy coding.

2.1.3 Distortion and Quality Metrics

Several distortion metrics are defined to quantify the degree to which the reconstructed image matches the original. Given an image of $M \times N$ pixels, the *Mean Square Error* (MSE) is defined as follows:

$$\text{MSE} = \frac{1}{M \times N} \sum_{i=1}^{M} \sum_{j=1}^{N} |P(i,j) - \hat{P}(i,j)|^2 \tag{2.1}$$

where $P(i,j)$ and $\hat{P}(i,j)$ represent the original pixel and the reconstructed pixel located at (i,j), respectively. Another frequently used

metric is the *Mean Absolute Difference* (MAD), defined as

$$MAD = \frac{1}{M \times N} \sum_{i=1}^{M} \sum_{j=1}^{N} |P(i,j) - \hat{P}(i,j)| \tag{2.2}$$

where $|x|$ represent the absolute value of x.

The most widely used distortion and quality metric in image and video coding is the *Peak Signal to Noise Ratio* (PSNR), which is defined as follows:

$$\text{PSNR} = 10 * log10(\frac{MAX_P^2}{MSE}) \tag{2.3}$$

where MAX_P is the maximum possible value of pixels in the original image. For 8-bit unsigned images, $MAX_P = 2^8 - 1 = 255$. Generally, increasing PSNR values indicates increasing reconstructed image fidelity.

As the quality of reconstructed images is eventually judged by end users, subjective distortion metrics that incorporate the human visual system (HVS) have been developed (47; 48; 49). The sensitivity of human visual system to different frequency and composition signals is different. By incorporating HVS models with objective metrics, more accurate quality metrics can be developed. The human visual system is very complex, which makes the development of such metrics difficult. Fortunately, objective metrics such as PSNR have been found to be consistent

with the observation of human being most of time (50). Therefore, PSNR will be used as the quality metric throughout this dissertation.

Distortion can be measured either in the transform or the spatial domain. Ideally we should calculate distortion in the spatial domain as this is directly linked to visual effect. However, very often we need to repeatedly estimate distortion and determine coding parameters while encoding is in progress. In this scenario it is not practical to perform an inverse transform and calculate the distortion in the spatial domain every time. Theoretically, as long as the transform is orthogonal, the distortion calculated in the transform and spatial domain should agree. The integer and floating point DWT used in JPEG2000 and the BPE coder, strictly speaking, are not orthogonal, but bi-orthogonal (38). However, the distortion discrepancy between them is negligible (21). Therefore, distortion reduction is calculated in the transform domain for rate-distortion optimization.

The PSNR performance may vary a lot using different coders. For example, SPIHT is reported to have 3dB gain over JPEG (36), which is a remarkable improvement from the perspective of modern coding. In many coders, the magnitude of PSNR improvement is much less than that. For instance, the PSNR of SPIHT using arithmetic coding (SPIHT-

AC) is within only 0.1-0.3dB different from that of JPEG2000 (21), and the BPE coder is within 1dB lower than JPEG2000 in strip coding mode (51).

2.2 Rate Control

Rate control is basically a rate distortion (R-D) optimization process used to achieve the best balance between rate and distortion. Basically, there are two major issues in rate control: one is to estimate quantization parameters such that the resulting bit rate approaches the target bit rate as closely as possible. The other issue is how to properly allocate bit rates to a number of coding units such that the overall distortion is minimized.

For lossy image and video coding, the lower bound for the rate at a given distortion is described by the R-D function (52). A typical plot of bit rate R versus distortion D, known as the rate distortion curve, is shown in Figure 2.6. Generally there is a tradeoff between the distortion and rate, i.e., distortion increases if rate decreases, and vice versa. Given a rate, the distortion-rate function, $D = f(R)$, function shows the minimum distortion that can be theoretically achieved.

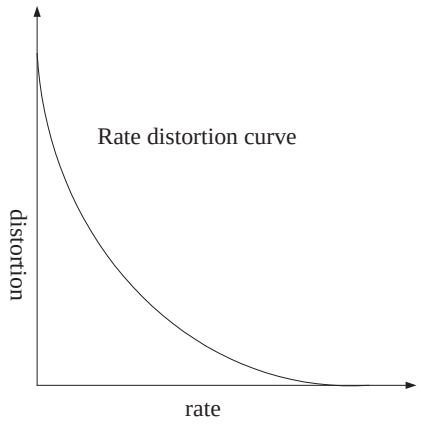
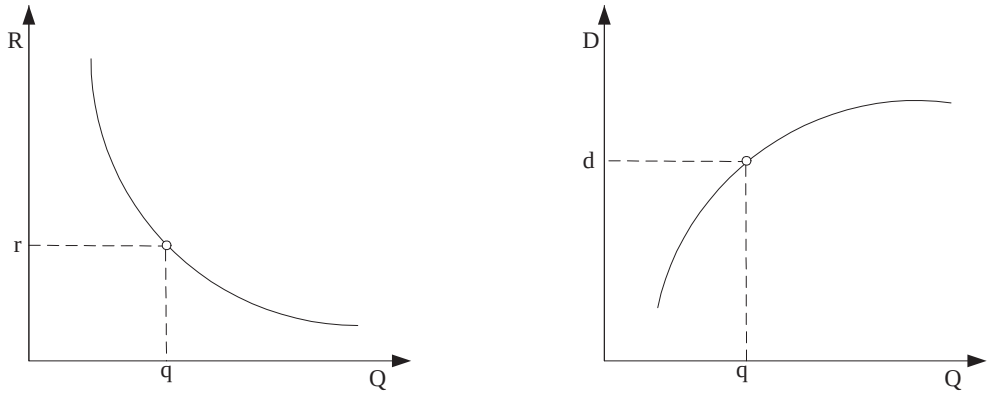


Figure 2.6: Rate distortion R-D curve

2.2.1 Rate Estimate

In lossy coding, the bit rate and distortion are considered as functions of quantization parameters, as shown in Figure 2.7. Basically, using a coarse quantizer generates lower bit rates and higher distortion than a fine quantizer. In video coding, quantization parameters are generally estimated using rate and quantization parameters models developed from past frames and the frame to be coded, as shown in Figure 2.8.

Rate estimates are used to determine quantization parameters such that the target bit rates are achieved. In practice, the discrepancy between target bit rates and achieved bit rates may cause unexpected coding behavior and should be minimized. Due to constraints on bandwidth or storage capacity, the extra bits may have to be discarded if they are



(a) Rate-Quantization Curve (b) Distortion-Quantization Curve

Figure 2.7: Relationship between rate, distortion, and quantization

beyond the bandwidth or storage capacity, resulting in uncontrolled distortion within an image. On the other hand, bit rates lower than the capacity waste coding resources and have to be avoided.

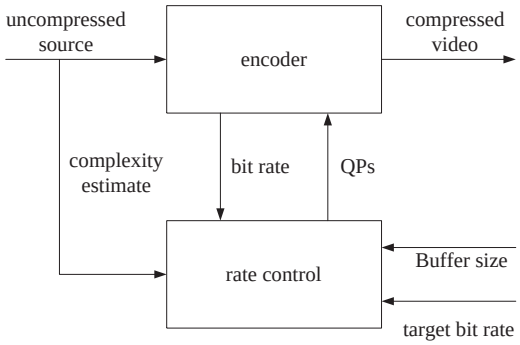


Figure 2.8: Block diagram of rate control

For many DWT based image coders that generate embedded bit streams, quantization indices are coded from highest bit plane to the lowest bit planes using bit scanning. Determining quantization parame-

ters is equivalent to determining the final bit plane of the quantization indices, and this can be viewed as inherent progressive uniform quantization.

2.2.2 Rate Allocation

Rate allocation is used to allocate the given bit resource to different coding units such that the overall distortion is minimized. Given a desired rate R for n coding units, we need to allocate $R(1)$,$R(2)$, $\cdots$, $R(n)$ bits to each coding unit, such that

$$D = \sum_{i=1}^{i=n} D(i) \tag{2.4}$$

is minimized, subject to the constraint:

$$\sum_{i=1}^{i=n} R(i) \leq R \tag{2.5}$$

where $D(i)$ is the distortion obtained from the ith coding unit with rate r_i, $i = 1, 2, \cdots, n$.

Scenes in natural video sequences may change rapidly and thus different frames may have different level of details and activity, as shown in Figure 2.9. A simple way for rate allocation is to fix quantization

parameters, or rate. However, this would lead to some practical coding issues.

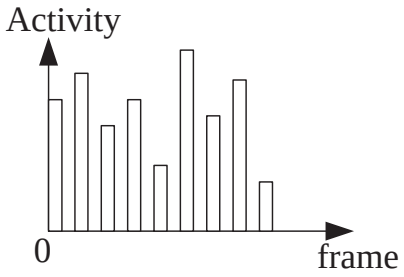


Figure 2.9: Activity of video frames

2.2.2.1 Fixed Quantization Parameters

If quantization parameters are fixed, the quality level of the compressed video is maintained. However, the rate may vary dramatically from one frame to another. For the frames that contain lots of details and motion, rate will go up quickly. While for the frames that do not contain many details and motion, rate will be very low. As a result, the overall bit rates are not predictable, which is not desirable in practice.

2.2.2.2 Fixed Bit Rate

A frame can be set to have fixed bit rate and variable quantization parameters. This may lead to video quality variations. A complex frame has to use coarser quantizers to achieve the target rate, resulting in loss

of details. On the other hand, a simple frame has to use finer quantizers to generate more bits in order to maintain the target rate. These would lead to inconsistent reconstructed video quality.

2.2.2.3 Adaptive Bit Rate Allocation

To avoid the problems with fixed bit rate and fixed quantization, adaptive rate allocation is necessary to allocate rate based on the content in video sequences. The basic principle is that fewer bits should be allocated to the frames with fewer details while more bits should be allocated to complex frames with more details and activity (53). This can maintain the video quality, while keeping the overall bit rate. Rate allocation is an essential problem in rate control and has been studied extensively for many years (26; 41; 44; 54; 55; 56).

A widely used method to address the problem of rate allocation is Lagrangian optimization (57). Assume that there are n coding units and for each coding unit, the rate-distortion function is known, i.e., $D(i) = f_i(R(i))$, where $i = 1, ..., n$. A Lagrangian cost function is built as follows:

$$J = \sum_{i=1}^{i=n} D(i) + \lambda \sum_{i=1}^{i=n} R(i) \tag{2.6}$$

Where λ is called the Lagrangian multiplier. Take partial derivatives

with respect to $R(i)$ and λ and set the resulting functions to zeros as follows:

$$\nabla(R, \lambda) = 0 \tag{2.7}$$

The resulting $R(i)$ obtained by solving the $n + 1$ equations turns out to be optimal in the sense that the overall distortion D cannot be reduced without increasing the overall rate R.

2.2.3 Rate Control for Video coding

Motion-compensated predictive video coding can be thought of as dependent coding due to high temporal dependencies and motion compensation. Given a picture or an image block, the available R-D points may depend on the R-D points of its reference blocks or pictures. If the reference blocks and pictures for motion compensation are coarsely quantized with low bit rates, motion estimation would be inaccurate and prediction error in motion compensation could be large, and therefore, lead to worse rate-distortion curves. If more bits are used for quantization of reference pictures, motion estimation is more accurate, but available bits for the future pictures are reduced. Rate control needs to handle this tradeoff to obtain the optimum coding performance. Lagrangian optimization based multi-level rate control algorithms have been proposed

in practice (53; 58; 59; 60).

Generally rate control algorithms are recommended in video coding standards, however, they are not a part of standard because rate control is a problem that depends on specific applications. The recommended rate control schemes do not aim to provide optimal solutions. The R-D model was not used in rate control for early video coding standards such as H.261 (13). H.261 (13) is mainly for low motion and stationary background video sequences. In its TM8 rate control scheme (61), the quantizer step size q does not incorporate the statistics of the video sequences and is expressed by

$$q = 2 \left\lfloor \frac{B}{200p} \right\rfloor \tag{2.8}$$

where B is the buffer level, p indicates the desired bit rates in terms of $p \times 62$kbits/s, which is the bandwidth of the ISDN network.

For the newer standards, such as H.263 (14) and H.264/AVC (62; 63), R-D models were built up and these allow more precise estimate of quantization parameters and rate control. The rate control algorithm in the JM reference software for H.264/AVC (64) consists of three levels rate control, group of pictures (GOP) level, picture level, and basic unit level. Figure 2.10 shows the abstract model of H.264 rate controller.

- GOP level rate control: GOP structure refers to an I-picture, followed by all the P and B pictures. The GOP rate control allocates the target number of bits and the QPs are initialized, based upon the bandwidth, frame rate, the number of frames, and the previous GOP.

- Picture level rate control: QPs are calculated based on a quadratic model as follows:

$$R = C1 * \frac{MAD}{QP} + C2 * \frac{MAD}{QP^2} \tag{2.9}$$

  where R is the desired bit rate, $C1$ and $C2$ are two model parameters derived from linear regression combined with the previous pictures and updated for each picture. The current MAD is predicted based on previous picture.

- Basic unit level rate control: similar to the picture level rate control except that the object is a basic unit, such as a set of macroblocks (MBs, a block of 16×16 pixels).

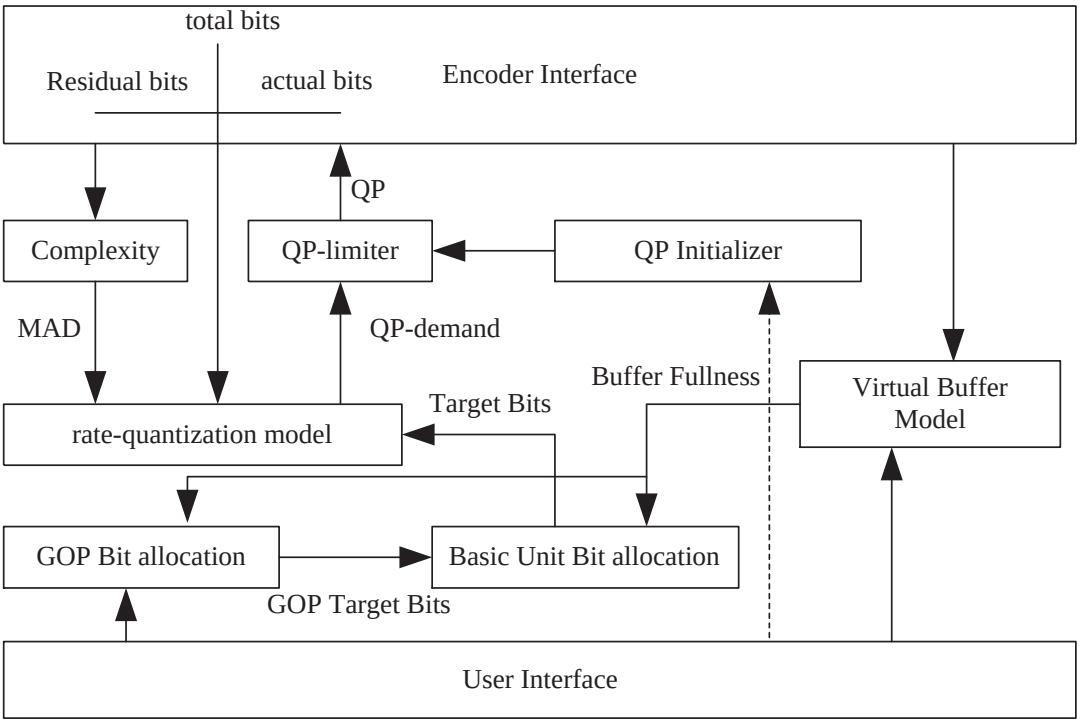


Figure 2.10: Elements of H.264 Rate Controller

2.2.4 Wavelet Video Coding and Rate Control

Wavelet video coders make use of the hierarchical pyramid structure of DWT and bit plane coding to generate an embedded bit streams. The embedded bit streams facilitate precise rate control up to bit level because the coders can truncate bitstream to ensure accurate rate control (42; 65; 66; 67; 68; 69; 70).

In (65) a modified layered zero coder is used as the basis of the video coder, which also incorporate the motion compensation algorithm employed by the MPEG-2. The rate distortion performance of the wavelet video coder is approximated by an exponentially decaying function and a Lagrangian optimization method is developed.

Lin and Gray proposed wavelet video coding with adaptive rate control based on the SPIHT algorithm in (42). They applied SPIHT locally to blocks that share similar characteristics rather than an entire image. Two kinds of optimization are performed. One is rate distortion optimization with Lagrangian optimization, and the other is dependent optimization based on an iterative method.

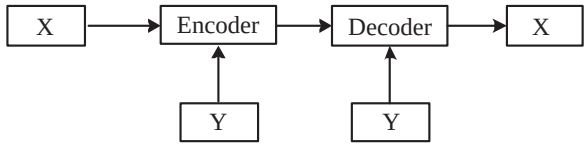
In (66), a unified mathematical model to quantify the relationship between the spatiotemporal wavelet decomposition structure, bit rate, and distortion is proposed for wavelet video coders with motion-compensated

temporal filtering (MCTF) structures. The subbands are coded independently and the bit rates of different subbands are optimally truncated to minimize the distortion under a bit-rate constrained scalable coding.

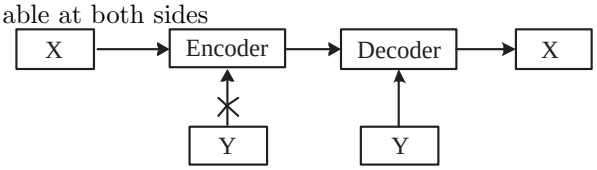
2.3 Distributed Source Coding

Established in 1970s, Slepian-Wolf coding (28) and Wyner-Ziv coding (29) theory states that separate encoding of two correlated sources X and Y can have the same coding efficiency as joint encoding of X and Y. As shown in Figure 2.11(a), traditional encoding of two correlated information sources requires knowledge of X and Y. In the system as shown in Figure 2.11(b), however, even if Y is not available at encoder when encoding X, the decoder is able to perfectly reconstruct X, as long as Y is available at decoder. This type of coding is often referred as *distributed source coding* (DSC).

Let $H(X)$ and $H(Y)$ be entropy of X and Y, respectively, and $H(X|Y)$ be the conditional entropy, the Slepian-Wolf theory proved that for X we can achieve a compression ratio of $H(X|Y)$, i.e., the same as the compression ratio of joint source coding with Y at encoder. Wyner-Ziv theory extended lossless Slepian-Wolf coding into lossy coding by introducing a quantizer, as shown in Figure 2.12.



(a) Traditional coder with side information Y available at both sides



(b) Slepian-Wolf coder with side information Y available only at decoder

Figure 2.11: Diagram of traditional coder and Slepian-Wolf coder. The information sequence Y can be viewed as a side information without error occurrences.

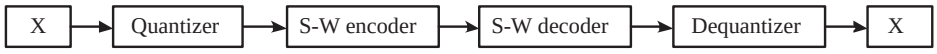


Figure 2.12: Wyner-Ziv coder: a Slepian-Wolf coder with a quantizer

The theories do not actually provide approaches to design such coders and much effort has gone into design issues. Distributed coding systems using the concepts of syndrome of block coding (71), powerful forward error correcting code such as Turbo code (1; 72; 73; 74; 75) and LDPC coding (76; 77; 78; 79) have been proposed in the last few years.

In traditional image and video coders, encoders have higher complexity while decoders are often simple. For instance, in H.263, MPEG-1 and MPEG-2 standards the encoders are five to ten times more complex than the decoders (80). Most approaches proposed under the framework of distributed source coding have simple encoders, but complicated de-

coders. This is particularly suitable for applications such as wireless sensor network, where the encoders cannot afford much computational load and complexity due to constraints of energy supply and memory and the decoders at the receiver stations do not have such restrictions.

2.4 Robust Video Coding

Robustness of video coding is a very practical and important issue when the compressed video sequences are transmitted over error-prone channels. Video coding can be made more robust by a number of techniques as follows:

- Forward error correction (FEC): FEC codes such as block codes and correlation codes (81; 82; 83; 84) can be applied to video for transmission to combat the effects of bit errors. This is probably the simplest and most straightforward way to enhance the robustness.

- Unequal error protection (UEP): UEP is one of the most important tools in video communication systems over error-prone channels. In this scheme, the codewords that are more importance to visual quality are better protected than the others. Therefore the codewords need to be categorized according to their importance to visual qual-

ity and sensitivity to errors (85; 86; 87). One method is to analyze the video frames and then partition images to different regions. The regions of interest are protected using more bit resources in order to maintain a good image quality (88).

- Robust source code: variable length codes (VLCs) are widely used in image and video coding. However, they may lose synchronization if bit errors occur in the bitstream. Restart markers or synchronizing symbols can be added to the bitstream to help the decoding synchronized. Alternatively, robust source code such as the reversible variable length code (RVLC) (89; 90; 91; 92; 93) and T-codes (94; 95; 96; 97) can be employed to correct certain type of bit errors.

- Joint source channel coding (JSCC): source code can be optimized based on the channel characteristics. This usually involves modeling of the effects of channel errors on the decoded video quality, and a joint optimization between source and channel coding (98; 99; 100; 101; 102; 103; 104; 105; 106; 107).

- Error concealment (108). Error concealment is a method to recover or conceal the loss information due to the transmission errors. Spa-

tial or temporal interpolation and advanced decoding strategies can be employed to estimate and refine the corrupted data set. Using error concealment can improve the reconstructed video quality at the decoder.

- Distributed video coding: in addition to the main bit stream for each frame, an extra bitstream that is generated based on the principle of distributed source coding is sent to the decoder to assist the decoding (1; 73; 77; 87; 109; 110). By providing additional correlation information, error propagation can be limited and robustness can be improved significantly.

- Multiple description coding: given a data set, two or more coding descriptions are created, each of which may have low level resolution. Each version may be transmitted over different channels to overcome the possible entire information loss if there is only one channel. Combining with all these low resolution descriptions, the decoder can reconstruct the data set with high resolution. So losing one description will not severely degrade the quality (111; 112).

2.5 Bit Plane Encoder (BPE)

The BPE coder is a wavelet based image coder that was standardized by CCSDS in 2007 (23; 51; 113). It is intended to be used for on-board spacecraft, and its low memory requirement and complexity makes high-speed onboard hardware implementation feasible.

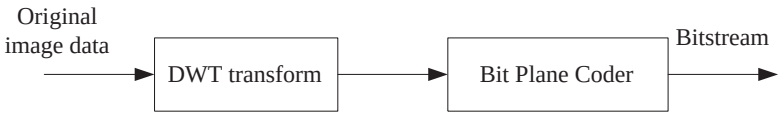


Figure 2.13: Encoder of the BPE coder

In this standard two types of DWTs were adopted: one is 9/7 floating point DWT (114) and the other is 9/7 integer DWT (115). The floating point DWT has excellent compression performance, however, it does not provide lossless compression. On the other hand, the integer DWT has inferior compression performance but it provides lossless compression.

In the BPE coder, an image is first decomposed using a 3-level decomposition as described in Subsection 2.1.1, and ten subbands are generated. The resulting DWT coefficients are compressed by the "Bit Plane Coder" in Figure 2.13. Figure 2.14 shows the flowchart and bitstream component of the BPE coder. We briefly review the key steps. The bit scanning process of AC coefficients is described by a series of words using

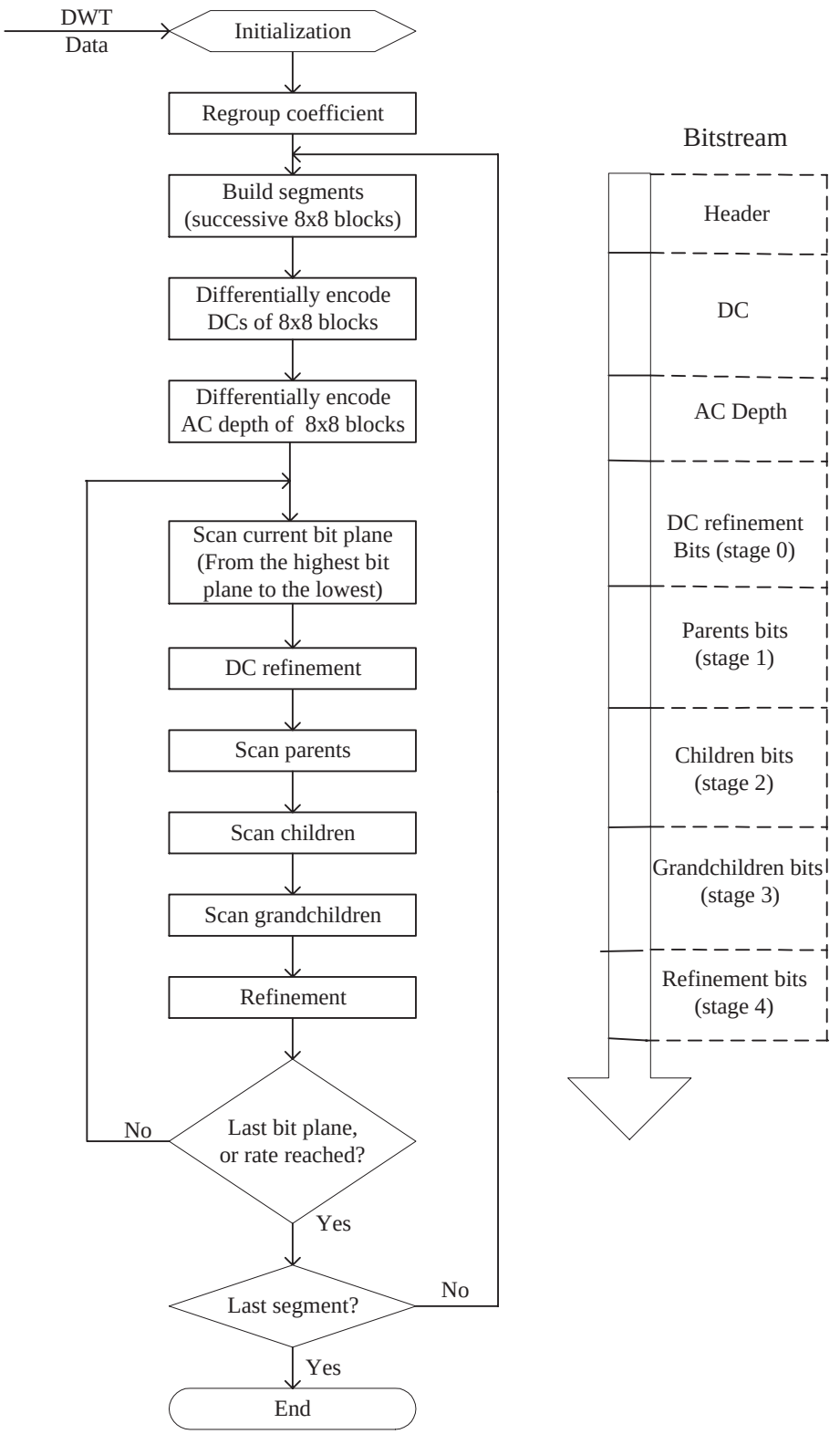


Figure 2.14: Flowchart of the bit plane encoder

coding structure to exploit dependency between coefficients in a block. (More information on the standard is available in the CCSDS blue book (23) and the CCSDS green book (51)).

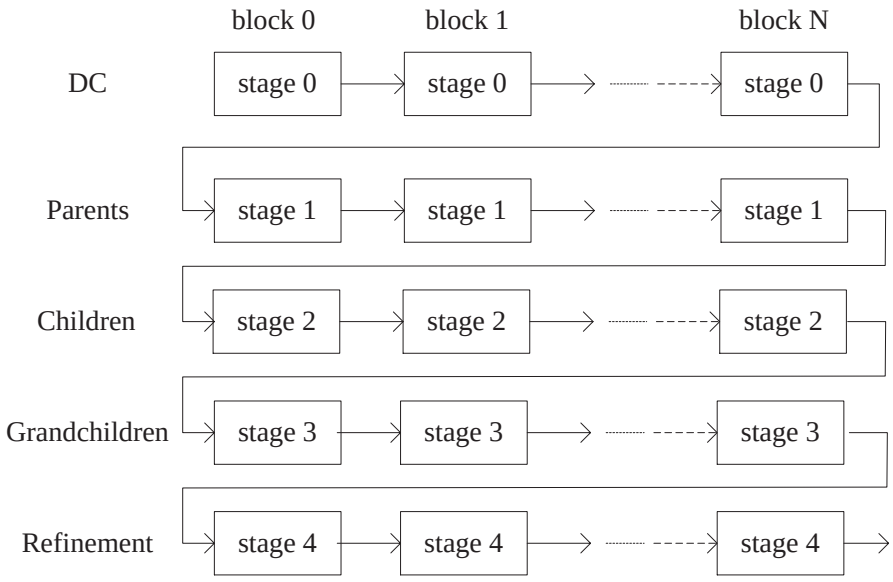
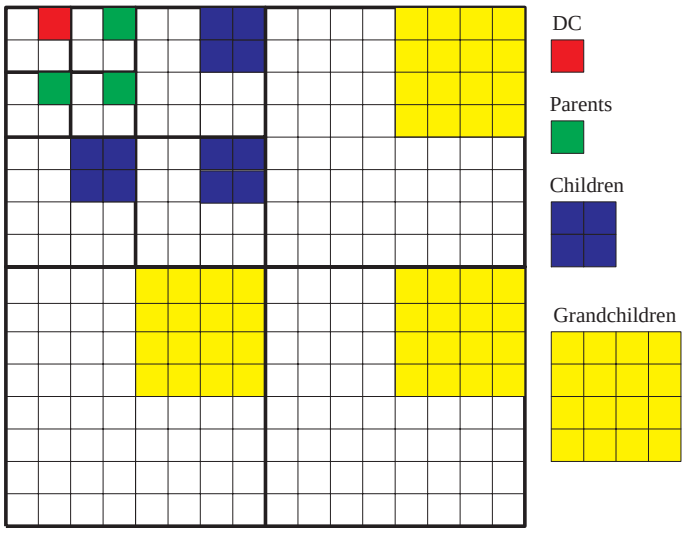
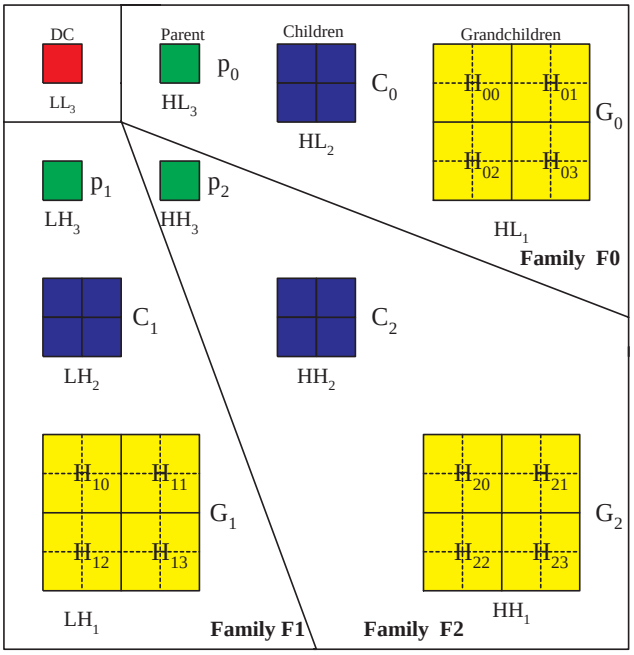


Figure 2.15: Structure of an encoded bit plane

1. Regroup coefficients and produce coding blocks: the wavelet coefficients are processed in groups of 8×8 coefficients, which are referred to as blocks. Each block consists of a single DC coefficient and 63 AC coefficients, as shown in Figure 2.16(a). Figure 2.16(b) illustrates a single block of coefficients and the family structure. The AC coefficients in a block are classified to three families, $F0$, $F1$ and $F2$. Each family $F(i)$ in the block has one parent coefficient p_i, a set C_i of four children coefficients, and a set G_i of sixteen grand-



(a) 2D DWT on image and the block structure



(b) DC, parent, children, and grandchildren in 8×8 block

Figure 2.16: DWT coefficients and how the blocks are reorganized

children coefficients. A group of N consecutive blocks are defined as a *segment*, where N is specified in the segment header. DC coefficients are represented using 2's complement representation. For each segment, **BitDepthDC** is defined as the maximum number of bits required to represent the DC value over all DC coefficients in the segment. ACdepth(m) is defined as the maximum number of bits required to represent the magnitude of any AC coefficient in the mth block. In the **Segment**, **BitDepthAC** is defined as the maximum value of ACdepth(m), where $m = 1, 2, ..., N$.

2. Segment header: every segment has a header to specify coding parameters used for the current segment. The header can have up to four parts. The first part is mandatory and contains information such as segment counter, BitDepthDC, and BitDepthAC, as shown in Table 2.1. The remaining three parts are optional and the last three bits of the first part of the header indicate whether the optional header parts are included. Note that the second part of the header, as shown in Table 2.2, contains a parameter called SegByteLimit. This parameter is to specify the number of bytes allocated for current segment, which will be used for rate control in next chapters. The third and fourth part of the header spec-

ify parameters that are generally fixed for an entire image, such as the DWT type, image width, transform weighting factors, and bit depth of the original image.

Field	bits	Description
StartImgFlag	1	Flags initial segment in an image
EndImgFlag	1	Flags final segment in an image
SegmentCount	8	Segment counter value
BitDepthDC	5	Number of bits needed to represent DC coefficients
BitDepthAC	5	Number of bits needed to represent AC coefficients
Part2Flag	1	Indicates presence of Part 2 header
Part3Flag	1	Indicates presence of Part 3 header
Part4Flag	1	Indicates presence of Part 4 header

Table 2.1: First part of the header

Field	bits	Description
SegByteLimit	27	Maximum number of compressed bytes in a segment.
DCStop	1	Indicates whether compressed output stops after coding of quantized DC coefficients
BitPlaneStop	5	indicates limit on coding of DWT coefficient bit planes.
StageStop	2	indicates the stage at which the coding stops.
UseFill	1	Specifies whether fill bits will be used to produce SegByteLimit bytes in each segment.

Table 2.2: Second part of the header

3. Coding of DC coefficients: the DC coefficient of each block in a segment is quantized, resulting in N quantized DC coefficients. The N quantized DC coefficients are then differentially coded using variable length code.

4. $ACdepth(m)$, where $m = 1, 2, ..., N$, are coded using the same differential and variable length coding procedure as the coding of quantized DC values.

5. Bit planes of coefficients: an AC coefficient is represented using the binary representation of the magnitude of the coefficient, along with a sign bit. In a segment the bit planes of the binary representation are encoded successively from the most-significant bit plane (MSB) to the least-significant bit plane (LSB). Within a bit plane, the coding of coefficients is performed in several stages, as shown in Figure 2.15. Stage 0 is used to refine the DC coefficients and it presents only when the current bit plane is below a threshold. The threshold is derived based on the quantization level of DC coefficients and the value range of AC coefficients. To better describe the bit scanning process, the AC coefficients are classified into lists based on their locations: the list of parents in the block is defined as $P = \{p0, p1, p2\}$; the list of descendants in family i, denoted D_i, is defined as $D_i = \{C_i, G_i\}$; the list of descendants in a block, denoted B, is defined as $B = \{D_0, D_1, D_2\}$. p_i, C_i, and G_i are illustrated in Figure 2.16(b).

- At stage 0, the bth most significant bit of the 2's-complement representation of the DC coefficient is output directly, where b is the current bit plane. This is to provide further DC coefficient resolution on the basis of the DCs that have been differentially coded.

- At stage 1 the parent AC coefficients in the segment are coded. Two words, $typesb[P]$ and $signsb[P]$, are defined as follows.

  – $typesb[P]$: denotes the binary word consisting of the bth magnitude bit of each parent coefficients.

  – $signsb[P]$: denotes the binary word consisting of the sign bit of each parent coefficients.

- Coding of children coefficients at stage 2. It contains a few words as follows:

  – $tranB$: transition bit. If the descendent coefficients in B becomes significant at this bit plane, $tranB = 1$. By default it is equal to 0.

  – $tranD$: indicates if the descendent coefficients in D becomes significant.

  – $typesb[Ci]$ and $signsb(Ci)$: $typesb[Ci]$ denotes the binary

word consisting of the bth magnitude bit of the coefficients in C_i, and $signsb(Ci)$ represents the sign of the coefficients in C_i.

- The grandchildren coefficients at stage 3. If $tranB = 0$, then stage 3 is unnecessary. Otherwise stage 3 consists of:
  - $tranG$: transition bit. It indicates if the descendent coefficients in G becomes significant.
  - $tranH_i$: transition bit. It indicates if the descendent coefficients in H_i becomes significant.
  - $typesb[H_{ij}]$ and $signsb[H_{ij}]$ $typesb[H_{ij}]$ denotes the binary word consisting of the bth magnitude bit of the coefficients in H_{ij}, and $signsb(H_{ij})$ represents the sign of the coefficients in H_{ij}.
- Mapping and entropy coding: the words generated from above procedure are mapped to integer values referred to as symbols, based on some pre-defined tables. The symbols are encoded using Golomb-Rice code (116; 117).

6. AC refinement: if an AC coefficient has been marked as significant in a higher bit plane, its bit at current bth bit plane is output

directly without compression.

We will illustrate the AC scan process using a simplified example. Figure 2.17 shows 4×4 coefficients from the DWT, where we assume all grandchildren coefficients are zeros and omitted. The coefficients in each list are as follows:

- $P = \{-6, 10, 5\}$
- $C_0 = \{2, 5, 2, 0\}$
- $C_1 = \{3, -5, 0, 0\}$
- $C_2 = \{-1, 3, 3, 0\}$

We can see that the maximum magnitude of AC coefficients is $p_1 = 10$ (1010 in binary). Therefore, the BPE coder scans from the fourth bit plane (this information is coded and transmitted in $ACdepth(m)$).

- Parent coefficients: at fourth bit plane, $p_1 = 10$ is significant, while $p_0 = -6$ (0110 in binary) and $p_2 = 5$ (0101 in binary) are not significant. Therefore $typesb = 010$ (in binary) and $signsb[P] = 0$, as $p_1 > 0$. Note that $signsb[P]$ is only present for significant coefficients.

- Children coefficients: as all the descendent coefficients are insignificant at fourth bit plane, $tranB = 0$. The bit scan ends for fourth bit plane.

Now the BPE coder proceeds to the third bit plane.

- Parent coefficients: As p_1 has been marked significant at fourth bit plane, it is omitted at the third bit plane at this stage. As $p_0 = -6$ and $p_2 = 5$ turn to significant at this bit plane, $typesb = 11$ (in binary). $signsb[P] = 10$ (in binary) as $p_0 < 0$ and $p_2 > 0$.

- Children coefficients: $tranB = 1$, as two children coefficients become significant. $tranD = 110$ (in binary), as both C_0 and C_1 contain one coefficient that become significant while C_2 has no significant coefficients. The BPE coder now scans the children coefficients in C_0 and C_1. $typesb[C_0] = 0100$ (in binary) and $signsb(C_0) = 0$, as $C_0 = \{2, 5, 2, 0\}$. Similarly, $typesb[C_1] = 0100$ (in binary) and $signsb(C_0) = 1$ (in binary).

- AC refinement: As P_1 has been marked significant at fourth bit plane, its bit at the third bit plane, 1, is sent to bitstream without coding. Coding of the third bit plane ends.

Now the BPE coder proceeds to the second bit plane.

- Parent coefficients: As p_0, p_1 and p_2 have been marked significant at the third and fourth bit planes, no coding is necessary at this stage.

- Children coefficients: $tranB$ is omitted. The BPE coder always scans the coefficients once $tranB$ has been set to 1 at higher bit planes. $tranD = 1$ (in binary), as C_2 contain one significant coefficient now. The BPE coder needs to scan all the children coefficients in C_0, C_1, and C_2. We can see $typesb[C_0] = 110$ (in binary) and $signsb(C_0) = 00$ (in binary). Similarly, $typesb[C_1] = 100$ (in binary) and $signsb(C_0) = 0$ (in binary). For C_2, $typesb[C_2] = 0110$ (in binary) and $signsb(C_2) = 00$ (in binary).

- AC refinement: As P has been marked significant at fourth bit plane, three bits at the second bit plane, 110 (in binary), are sent to bitstream without coding. 00 are sent to refine the two children coefficients (5,5) that had been coded at the third bit plane.

The BPE coder now proceeds to the last bit plane and all coefficients can be coded. Note that the words such as $typesb[C_i]$ and $tranD$ will be not be directly sent to bitstream. They are first mapped to symbols based on tables and the symbols are coded using variable length code.

13	-6	2	5
10	5	2	0
3	-5	-1	3
0	0	3	0

Figure 2.17: 4x4 coefficients from DWT

Within each segment, the bit plane scan can be regarded as a granular quantization refinement process and the scanning stops once the target bit rate is achieved. Thus the target bit rate can always be reached by a simple bitstream truncation. While for coding algorithms such as JPEG, JPEG2000, and MPEGs, the quantization parameters generally need to be determined in advance.

2.6 Conclusions

This chapter has provided a brief review of a number of topics in the field of image and video coding. We have discussed many interesting topics in image and video coding, including transform coding, rate control, motion compensation, distributed source coding, and the new wavelet based image standard BPE coder. In the following chapters we discuss in more detail and present work to address the issues of rate control

for wavelet based image coding, region based image coding, and wavelet based video coding using distributed source coding.

Chapter 3

Rate Control for Wavelet Based Image Coding

Rate estimation and rate allocation are two fundamental problems for image and video coding and have significant impact on coding performance. As we know, for wavelet based bit plane scanning coding, the target bit rate can be easily achieved by truncating the bitstream once the target bit rate is obtained. In this scenario, we focus on the rate allocation problem for wavelet based image coding.

In Section 3.1 a post compression R-D optimization based rate allocation (the PCRD-Opt algorithm) is discussed. In Section 3.2 a rate control scheme combining the PCRD-Opt algorithm with ρ domain analysis of wavelet coefficient is proposed. The compression performance using different rate allocation schemes with BPE coder is discussed and compared in Section 3.3. The chapter is concluded in Section 3.4.

3.1 Rate Allocation Using Post Compression Rate-Distortion Optimization

The rate allocation problem is the allocation of bit resources to coding units such that the overall distortion is minimized. The coding unit can be different in different coders. For instance, in H.264 the minimum coding unit is called a macroblock, which consists of 16×16 pixels and has its own quantization parameters. For JPEG2000, the coding unit is defined as 64×64 coefficients in the subbands obtained from the DWT decomposition. Under the assumption that the distortion of each coding unit is additive if they are coded separately, the total distortion can be obtained by summing up the distortion of each coding unit as follows:

$$D = \sum_i D_i^{n_i} \tag{3.1}$$

where n_i is a feasible truncation point on rate distortion curve selected from a number of candidate points and $D_i^{n_i}$ is the corresponding distortion incurred if the bitstream is truncated at n_i in coding unit i. Let R be the total number of bits to be allocated and $R_i^{n_i}$ be the number of bits allocated to the ith coding unit at n_i, the sum of bits allocated to

each coding unit, i.e., $\sum_i R_i^{n_i}$, is subject to the constraint as follows:

$$\sum_i R_i^{n_i} \leq R \tag{3.2}$$

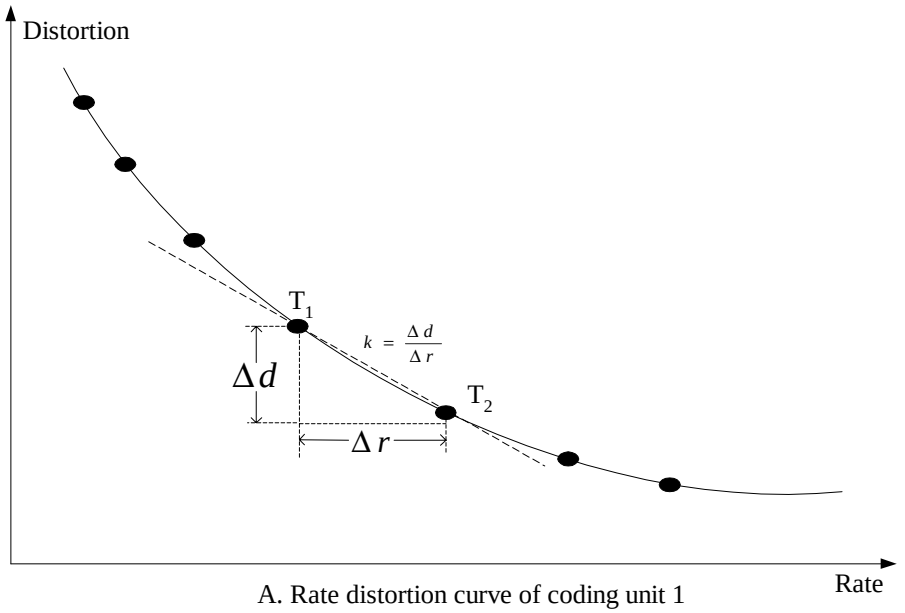
The optimum rate allocation is to determine a set of n_i such that the overall distortion D is minimized subject to the constraint condition in Equation 3.2.

This problem can be solved using the method of Lagrange multipliers as discussed in Section 2.2.2.3. The resulting λ can be interpreted as the slope of the rate distortion curve of each coding unit which minimizes the distortion as follows.

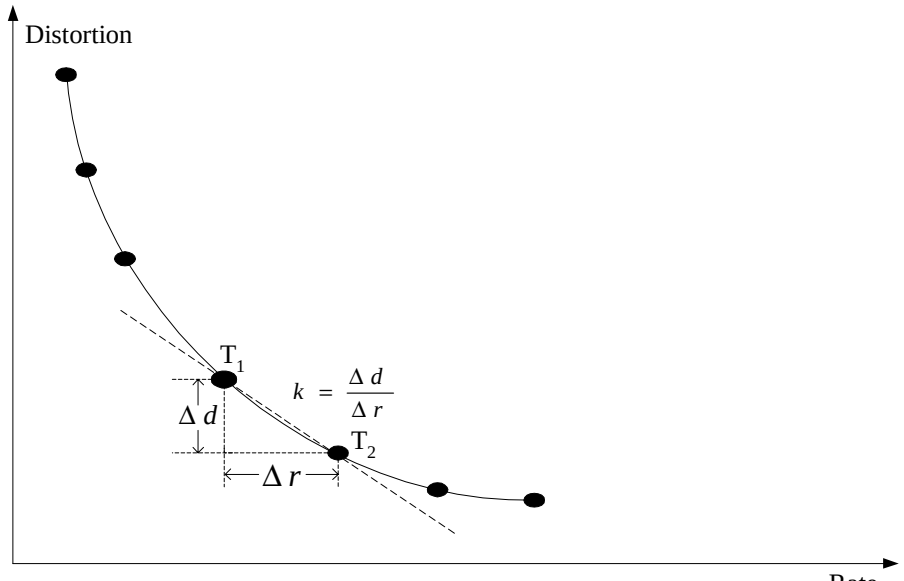
$$\lambda = \frac{\triangle D_i(n)}{\triangle R_i(n)} \tag{3.3}$$

It implies that the rate of distortion change with respect to bit rate is the same for the optimum point on each rate distortion curve. This is called the *Principle of Equal Slopes* (21). As shown in Figure 3.1, the two coding units have two different rate distortion curves while the resulting truncation points have the same slope.

Assume that all feasible truncation points on the R-D curve are available. An iterative search method could be used to find the truncation points that achieve the target rates and minimize the overall distortion



A. Rate distortion curve of coding unit 1



B. Rate distortion curve of coding unit 2

Figure 3.1: Rate distortion curve and its slope

as follows (21):

- *Iterative algorithm for post-compression rate optimization*
- *Step 1: Calculate the slope* l_i*, where* $i = 0, 1, 2, ..., n$ *at each candidate truncation point.*
- *Step 2: Find feasible truncation points. The feasible truncation points must satisfy the convexity condition, i.e., their slope must be less than the slopes of previous truncation points.*
- *Step 3: bisectional search to find the ideal* λ*. Takes the average of the smallest slope and the largest slope and then calculate a rate. If the rate is greater than the target rate, replace the smallest slope with the average slope, otherwise replace the largest slope with the average slope. This process iterates until the rate is reached, or the number of iterations exceeds the pre-defined threshold.*

To calculate the slope at each truncation point one needs the knowledge of distortion reduction and rates consumed from Equation 3.3 prior to the iterative method, and a considerable amount of computational resources are required. For instance, it turns out that two thirds of computational resources are used for the PCRD-Opt algorithm in JPEG2000

(21). Therefore, the PCRD-Opt algorithm may not be feasible for applications that impose strict complexity constraints.

3.2 Joint ρ Domain and PCRD-Opt Based Rate Allocation

In this section, an algorithm that combines the PCRD-Opt algorithm with the recently emerged ρ domain analysis is developed for rate allocation. A linear model based on coefficient analysis is developed, which is used to alleviate the severe computational burdens by the conventional PCRD-Opt. We will illustrate how this algorithm works using the BPE coder.

3.2.1 ρ Domain Analysis

As shown in Figure 2.7, rate R and quantization parameters QPs are generally not linearly related and it is difficult to express R in terms of QPs in closed-form formulas. However, it has been shown (24; 25; 26) that there exists a linear relationship between R and ρ in the transform domain for many image coders as follows:

$$R(\rho) = (1 - \rho)\theta \tag{3.4}$$

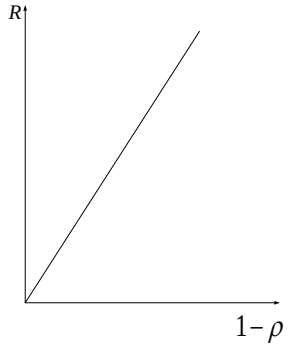


Figure 3.2: linear relationship between rate R and $1-\rho$

where θ is a constant that depends on the image and image coders and ρ is the ratio of the number of coefficients that are quantized to zero over the total number of coefficients. In extreme cases, assuming that all coefficients are quantized to 0, ρ becomes 1 and the resulting rate is $r = 0$.

This linear relationship has been shown to hold if the transform coefficients follow a Laplacian distribution or a generalized Gaussian distribution (24). It shows that generally for natural images the coefficients from the DCT and the wavelet transform fit well a Laplacian distribution or a Gaussian distribution (118; 119).

If θ in the linear model is available, one can either estimate the rate R for given quantization parameters QPs, or estimate quantization parameters QPs for the target rate R. θ can be obtained by checking available

points on the $R-\rho$ curve. The rates corresponding to other points can be obtained thereafter. ρ domain analysis has been shown to achieve excellent performance on rate estimation for video coders.

3.2.2 ρ domain analysis for the BPE coder

We intend to examine if ρ-domain analysis is valid for the BPE coder. If an accurate $R-\rho$ model is available, rate estimation and allocation could be facilitated. The model based on Equation 3.4 may not be directly applied to the BPE coder, as the BPE coder has header bits, DC, and AC depth prior to bit plane scanning and when $\rho = 1$, R is not equal to 0. Therefore, we introduce an offset γ to account for those bits as follows:

$$R(\rho) = (1-\rho)\theta + \gamma \tag{3.5}$$

The percentage of the zeros in each bit plane ρ_i and the number of bits r_i that has been used to encode the bit plane can be obtained, and for each bit plane, we have $r_i = (1-\rho_i)\theta + \gamma$. We use a first-order linear regression to obtain a set of parameter θ and γ such that the MSE between the rates obtained using this set of parameters and the actual rates is minimized.

In first experiment, we take an image as a single coding unit and

compress it using the BPE coder. We then visually examine if this linear model is valid for the BPE coder. It turns out that this model works very well. Figure 3.3 illustrates the actual $R-\rho$ curve and the one obtained using the linear regression on the Lenna image, and we can see that they appear to match very well.

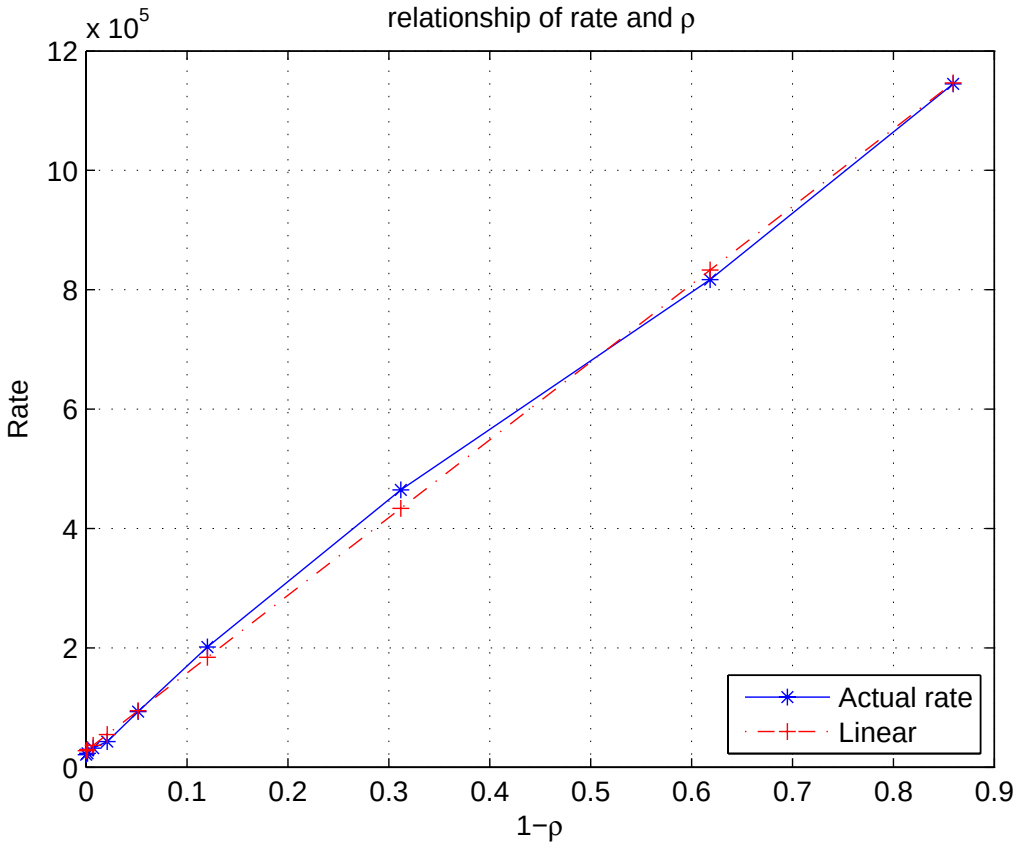


Figure 3.3: $R-\rho$ curve and curve fit for Lenna image, where the whole image is treated as a single coding unit and the linear regression method is used.

Secondly, we divide an image into a number of segments, where each segment consists of equal number of 8×8 blocks. Then each segment is

tested to see if this model is valid. It is found that the model works for small coding units as well. Figure 3.4 illustrated the $R-\rho$ curve for 16 equal-sized segments, where each segment consists of $\frac{512*512}{16} = 256$ 8×8 blocks.

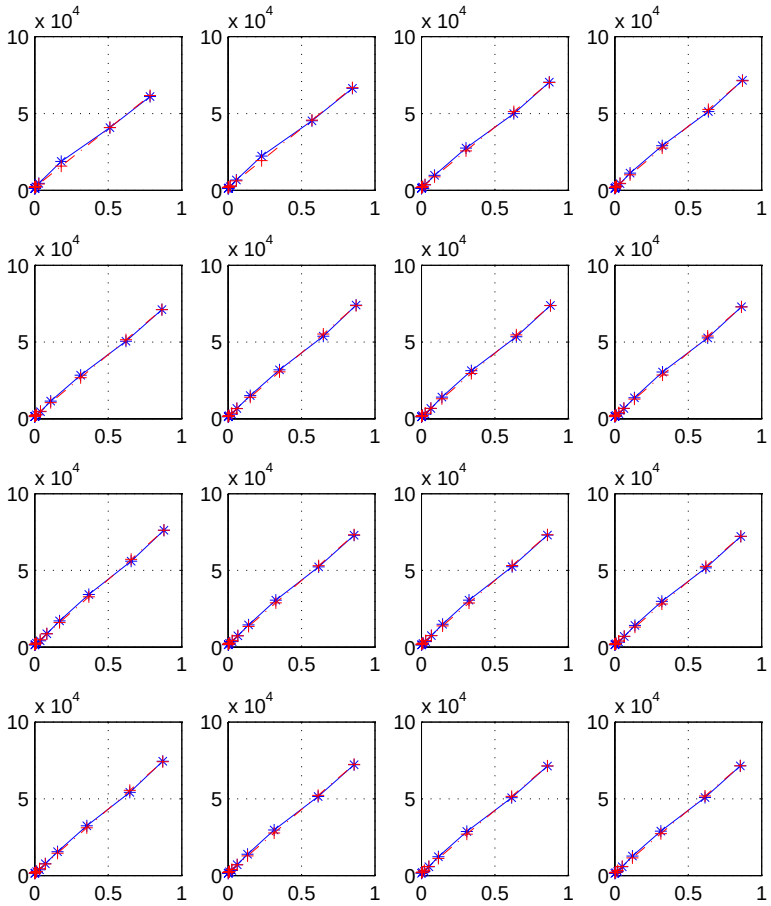


Figure 3.4: $R-\rho$ curve and curve fit of 16 segments of Lenna image, where the image is divided to 16 segments and each segment consists of 256 8×8 blocks.

3.2.3 Parameter Estimate of ρ domain for the BPE coder

As ρ domain analysis is valid for both cases, we will employ it to address the rate allocation problem. We attempt to estimate θ and γ in Equation 3.5 without completing the entire encoding process. After completing the encoding process at the first three highest bit planes (a high bit plane corresponds to large quantization step size and therefore, there are more zeros at the bit plane), an initial estimate of θ and γ can be obtained using the linear regression described above. The number of bits used to encode low bit planes (small quantization step size) can then be predicted using Equation 3.5.

We apply this method to predict the $R-\rho$ curve, and find the resulting predicted curve is away from the actual curve at low bit planes. In particular, the predicted rates tend to be constantly higher than the actual rates, and the prediction error decreases as the numbers of bit planes used for prediction increases, as illustrated in Figure 3.7. To explain this phenomena, we examine the theory of ρ domain analysis and it is found that the $R-\rho$ function is approximately linear provided that $1-\rho$ is close to zero. In other words, this one-order linearity is valid if most coefficients are quantized to zeros, i.e., ρ is close to 1. For the Laplacian source, the approximation in ρ domain analysis is expressed

as follows (24)

$$R(\rho) = 2\log_2 e \cdot (1-\rho) + O([1-\rho]^3) \tag{3.6}$$

At higher order bit planes, as $1-\rho$ is close to zero, $O([1-\rho]^3)$ can be omitted. However, when the encoding proceeds to lower order bit planes, the number of zeros decreases so that $1-\rho$ is no longer close to zero. Hence, we speculated that the absence of high order terms in $R(\rho)$ of Equation 3.7 causes the discrepancy. In order to improve the accuracy of the rate estimate, we take into account the high order terms that have been omitted using the model as follows:

$$R(\rho) = k_1 \cdot (1-\rho) + k_2 \cdot (1-\rho)^2 + k_3 \cdot (1-\rho)^3 \tag{3.7}$$

Using the data collected from the first four bit planes, k_1, k_2 and k_3 are obtained through the well known polynomial curve fitting method which minimizes the error in a least-squares sense. Figure 3.5 shows that the rate estimated verses $1-\rho$ using the new model parameters when treating the entire image as one coding unit, and Figure 3.6 shows the rate estimated verses $1-\rho$ for each coding unit.

We observe that the error between the actual rate and the estimate

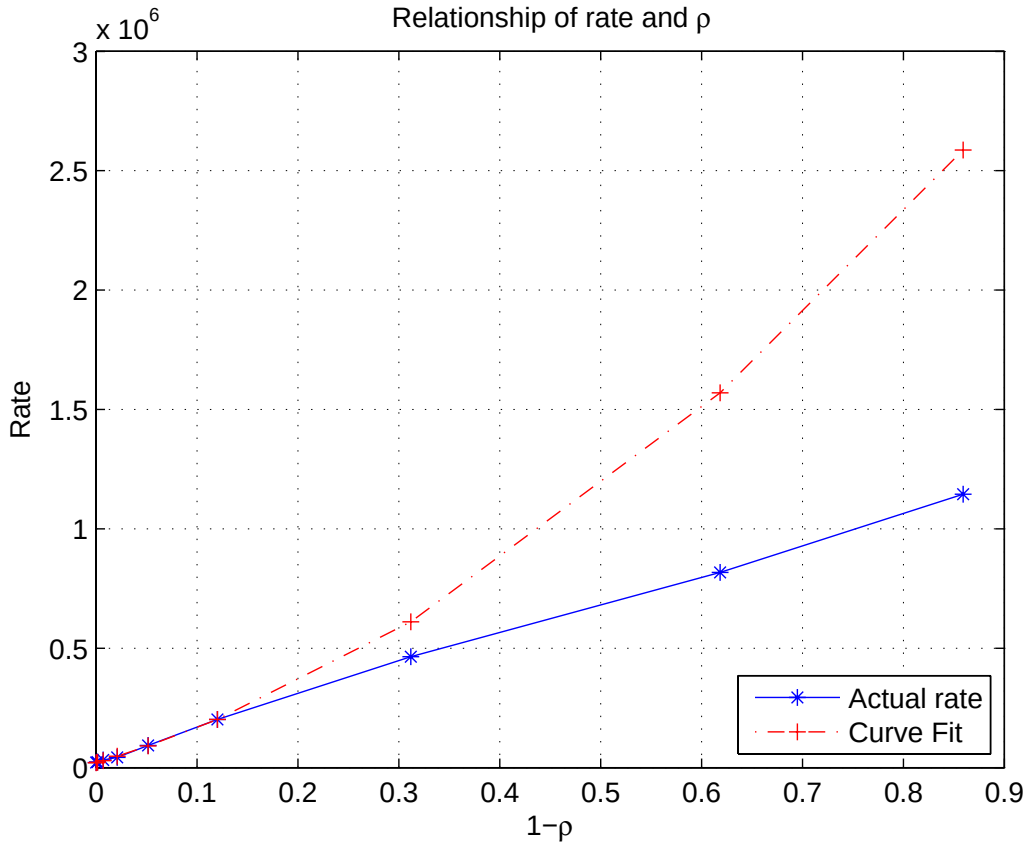


Figure 3.5: $R - \rho$ curves for the whole Lenna image. The curve marked as "Actual rate" is obtained using the actual number of bits obtained at corresponding bit planes; the curve marked "Curve Fit" is the one obtained using the model defined in Equation 3.7 and the information obtained from the first four bit planes.

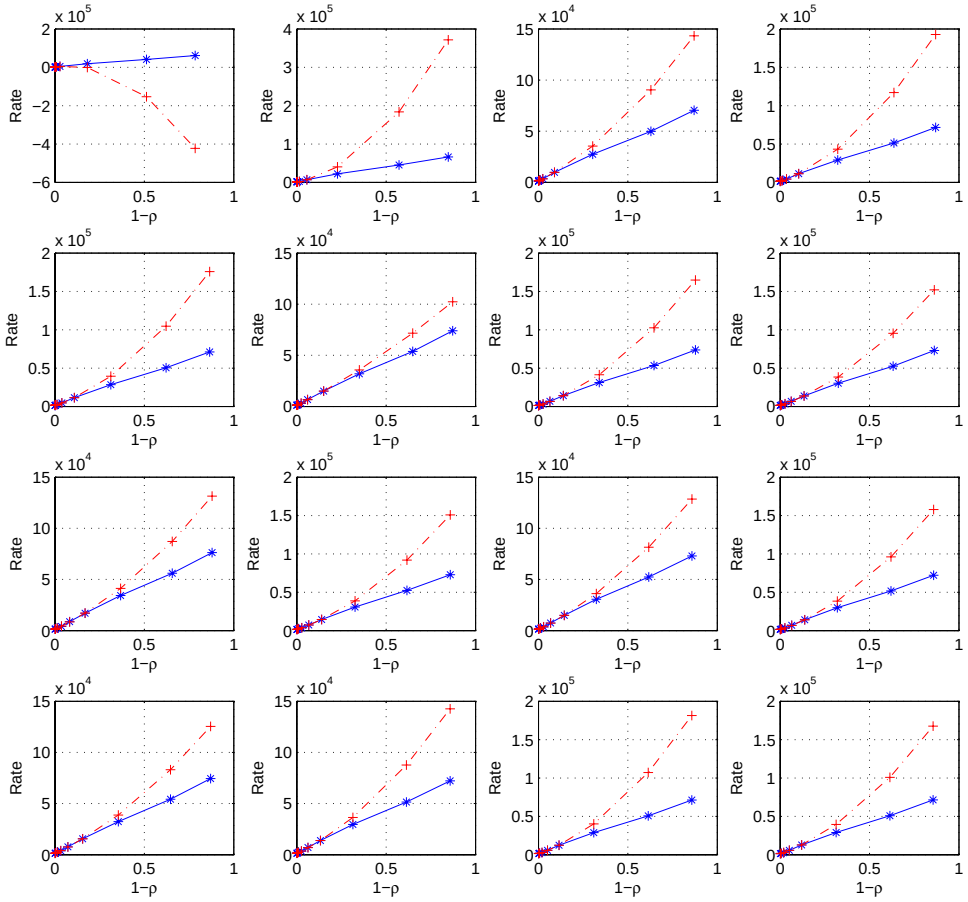


Figure 3.6: $R - \rho$ curves for the 16 coding units of Lenna image. The curve marked as "Actual rate" is obtained using the actual number of bits obtained at corresponding bit planes; the curve marked "Curve Fit" is the one obtained using the model defined in Equation. 3.7 and the information obtained from the first four bit planes.

obtained using Equation 3.7 is very large. This is surprising while understandable. For the higher order bit planes, ρ is close to zeros, and the $R-\rho$ curve has good linearity. Using those data as the basis to find the curve that fits those high bit planes, the high order terms do not have big impact on the curve. However, using the model to estimate the rate at low bit planes, a small variation or mismatch from the high order terms may have a significant impact on the rate estimation at lower bit planes. Because the non-linearity only appears at low bit planes where ρ is not close to zero, the curve derived from the higher order bit planes does not fit the lower order bit planes as expected.

To correct this, rather than using the high order model, we propose a two-stage linear model to estimate the rates at low bit rates. Our experiments show that at the lowest two bit planes, the predicted rates from ρ domain analysis by using the highest three bit planes are around 8% higher than the actual rates. Therefore, we modify the ρ model and estimate θ and γ using two steps as follows:

1. *The first linear regression is based on the first three bit planes to estimate θ and γ such that the predicted rates are closest to the target rates.*

2. *The predicted rate at lowest bit plane is scaled by* 8%*. This new data*

point, along with the available data points, are used to re-estimate a linear curve using the linear regression method. The resulting linear model is used to calculate the rate for the low bit planes.

The experiment shows that the predicted rates based on the modified linear model match the actual rates much better than the original model. Figure 3.7 illustrates the actual $R - \rho$ curve, the linear predicted $R - \rho$ curve, and the $R - \rho$ curve obtained using the modified 2-step model described as above, where the whole Lenna image is taken as a single coding unit. Figure 3.8 shows the three curves for the 16 coding units where each one contains 256 consecutive 8×8 coding blocks.

Note that for most segments the modified model works very well, except that for the first segment there is still a gap between the predicted and the actual rate. By examining the region we find that it contains little activity and the coefficients cannot be strictly modeled as Gaussian or Laplacian.

3.3 Experiment and Results

In this section, the PCRD-Opt algorithm and the proposed algorithm are applied to the BPE coder to demonstrate the performance. Using

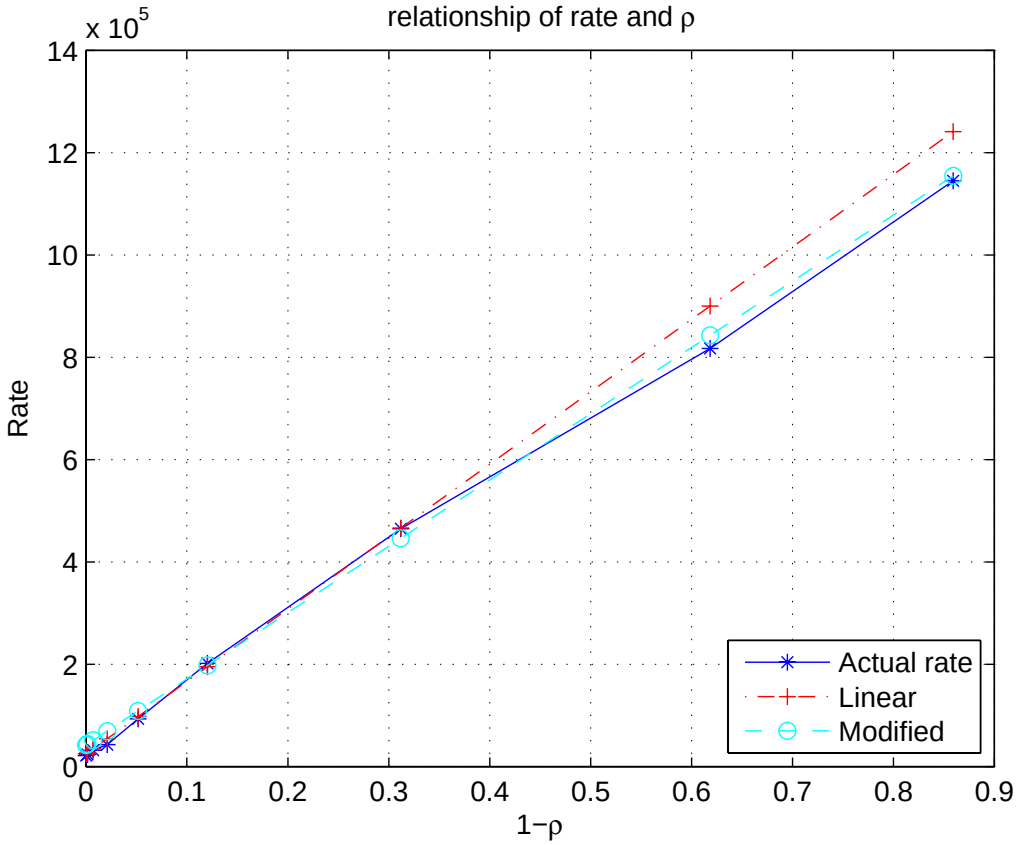


Figure 3.7: $R - \rho$ curves for the whole Lenna image. The curve marked as "Actual rate" is obtained using the actual number of bits obtained at corresponding bit planes; the curve marked "Linear" is the one obtained using the linear model and the encoding information of the first three bit planes; "Modified" is the one obtained using the modified 2-step estimation with linear regression.

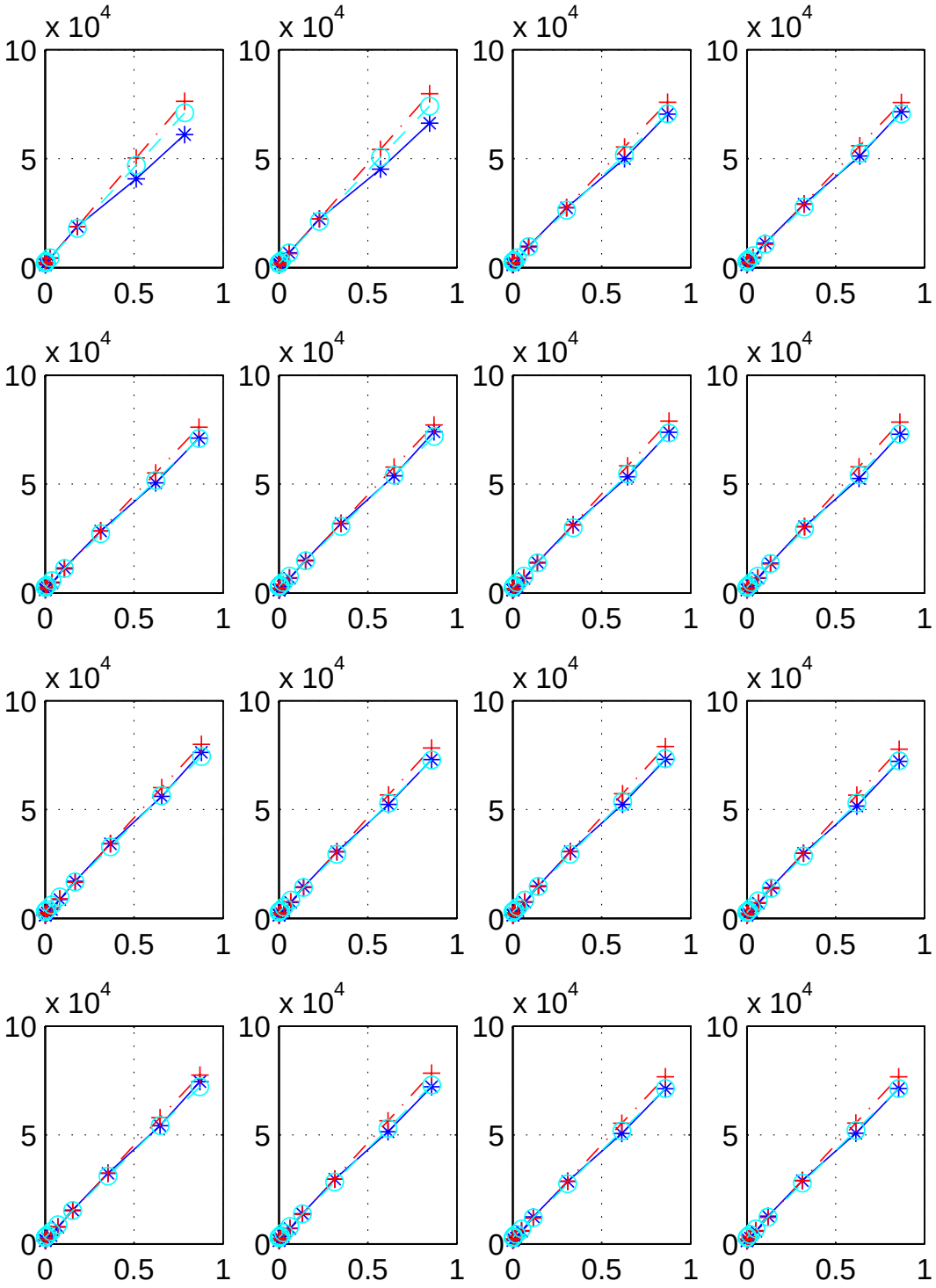


Figure 3.8: $R-\rho$ curves for the Lenna image, where the image is divided to 16 segments. "Actual rate", "Linear", "Modified" are interpreted in Figure 3.7

the model developed in Section 3.2, the number of bits used for coding bit planes can be estimated, then we can efficiently apply the principle of the PCRD-Opt algorithm for rate allocation.

3.3.1 The BPE coder with PCRD-Opt (PCRD-Opt BPE) Algorithm

The PCRD-Opt algorithm is adapted to apply in the BPE coder. We first classify the coding units and then define the methods to measure distortion reduction, and then report the coding performance.

3.3.1.1 Coding Units and Truncation Points

In the BPE coder as the coding of each segment is totally independent, we take each segment as the basic coding unit. We record the coding states as follows to be feasible truncation points:

- Coding of DC coefficients;
- DC refinement (stage 0);
- Three stages coding of AC coefficient (stage 0-3);
- AC refinement (stage 4).

The coding of AC depth after the coding of DC coefficients does not directly contribute to the distortion reduction because this information only assist the subsequent coding of AC coefficients. Therefore it is excluded from the candidate truncation points. For each segment, we record the distortion reduction and bits consumed.

3.3.1.2 Distortion Reduction

We need to record the distortion reduction for bit plane scanning. One widely used metric to measure distortion in image coding is MSE as defined in Equation 2.1. The distortion reduction after decoding of a bit depends on the reconstruction process at the decoder. For example, for the floating point DWT, assuming that the decoder decodes the bit plane q ($q = 1, 2..., q_{max}$, the lowest bit plane is 1), the distortion reduction by coding the pixel is given by

$$D_r = (2^q/2 - 2^{q-1}/2)^2 + 2^q * b_q \tag{3.8}$$

where b_q is the bit plane value of the pixel.

Given a coefficient x, assuming that a coefficient becomes significant at bit plane b_s at decoder, and is decoded as $\hat{x}_b$, the MSE reduction is

given by

$$D_{MSE} = x^2 - (x - \hat{x}_b)^2 = \hat{x}_b * (2x - \hat{x}_b) \tag{3.9}$$

Once a coefficient becomes significant, in the next lower bit plane $b-1$, its rate reduction is given by

$$D_{MSE} = (x - \hat{x}_b)^2 - (x - \hat{x}_{b-1})^2 = (\hat{x}_{b-1} - \hat{x}_b) * (2x - \hat{x}_b - \hat{x}_{b-1}) \tag{3.10}$$

where $\hat{x}_{b-1}$ represents the reconstructed x using the bit planes higher than $b-1$th bit plane. Note that for floating point DWT and integer DWT, the reconstruction of $\hat{x}$ is slightly different due to the different reconstruction consideration (51).

3.3.1.3 Experiment Results

Since JPEG2000 is also based on the wavelet transform and bit plane scanning, it is interesting to compare the compression performance of BPE with JPEG2000. The reference software we used for JPEG2000 testing is Jasper (120). In JPEG2000, every 64×64 coefficients are grouped into a coding unit for post-optimization. JPEG2000 can use different levels of DWT decomposition, and by default the level is set to five. However the BPE coder employs 3-level DWT decomposition. To make the comparison more fair, we use 3-level DWT decomposition for

JPEG2000.

bpp	Seg	64	128	256	512	1024	JP2k-3	JP2k-4	JP2k-5
0.2	VBR	30.83	31.05	31.2	31.23	31.31	31.56	32.27	32.2
	CBR	30.34	30.64	30.82	30.84	30.93			
0.4	VBR	34.69	34.73	34.99	34.95	34.86	35.08	35.29	35.31
	CBR	34.12	34.25	34.35	34.36	34.34			
0.6	VBR	36.58	36.59	36.59	36.58	36.67	36.93	37.03	37.11
	CBR	36.27	36.35	36.42	36.41	36.5			
0.8	VBR	37.92	37.94	38.05	38.14	38.14	38.26	38.4	38.36
	CBR	37.01	37.55	37.57	37.61	37.64			
1	VBR	38.99	39.01	39.16	39.02	39.02	39.16	39.28	39.29
	CBR	38.77	38.83	38.92	38.93	38.94			
2	VBR	43.05	43.08	43.17	43.09	43.1	43.42	43.36	43.38
	CBR	42.87	42.99	43.09	43.09	43.09			
3	VBR	47.17	47.18	47.21	47.2	47.28	47.33	47.42	47.42
	CBR	46.95	46.96	46.99	46.99	46.99			

Table 3.1: PSNR (in dB) of CR-BPE (CBR), the PCRD-Opt BPE (VBR) and JPEG2000, where JP2k-3, JP2k-4, and JP2k-5 represent the JPEG2000 using 3, 4, 5-level decomposition respectively

Table 3.1 shows the PSNR performance of the CR-BPE, the PCRD-Opt BPE, and the JPEG2000 for integer for the Lenna image. The results show that the average PSNR performance of the PCRD-Opt BPE is consistently better than the CR-BPE allocation, especially at low to middle bit rates. For example, when bit rate is 0.8bits/pixel, the gain of the integer DWT based PCRD-Opt BPE is around 0.55dB when the number of blocks in each segment is set to 512. This improvement is attributed to the fact that the PCRD-Opt algorithm has optimized the rate allocation such that the distortion is minimized. As the bit rates

increase, the gain decreases while it is still better than the original BPE.

In addition, PSNR performance with different number of blocks in each segment is tested. The number of blocks in each segment is set to 64, 128, 256, 512, and 1024 blocks, respectively. And as it increases from 64 to 256, the PSNR performance improves slightly. This is reasonable because the header of the segment is fixed regardless of the size of segments, and the percentage of header bits in small segments with respect to the total number of bits in the segment is higher than that in large segments. Therefore, relatively more bits are allocated to coding as the size of segment is increased. More specifically, the header of the first segment needs 152 bits, including part 4, which represents the unchangeable information throughout the encoding process. The headers of the subsequent segments cost 88 bits, representing the first three header parts.

As the number of blocks in each segment increases from 512 to 1024, the PSNR performance, however, does not monotonically improve. The PSNR drops at some bit rates for large segments. Figure 3.9 shows the PSNR performance when the number of segments is set to 16, i.e., each segment contains $\frac{512\times512}{64*16} = 256$ 8×8 blocks for an image with resolution of 512×512 pixels. By examining how the PCRD-Opt algorithm works

in the BPE coder, it is found that when the number of coding units is small, the discrepancy between the target bit rate and the resulting bit rate increases. This is not surprising as many truncation points do not satisfy the convexity condition and are ruled out from candidacy. In extreme cases when there is only one coding unit, this is the same as coding with constant rate and, therefore, it is certainly worse than the case that several coding units are presented. To alleviate the discrepancy, the resulting bit rates in all headers are scaled proportionally such that the desired rates are exactly the same as the actual rates. To simplify the discussion, the number of segments is set to 16 hereafter, unless otherwise specified.

For JPEG2000, as we can see, more levels of DWT result in trivial coding gain, as has been confirmed in (51). While the PSNR of the PCRD-Opt BPE coder is still below that of JPEG2000, the gap between the coders has decreased dramatically.

3.3.2 Joint ρ Domain and PCRD-Opt Rate Allocation

For rate estimate, ρ domain analysis enables accurate prediction of the resulting bit rate at bit planes without completing the coding and so the truncation points can be preset at the desired bit planes.

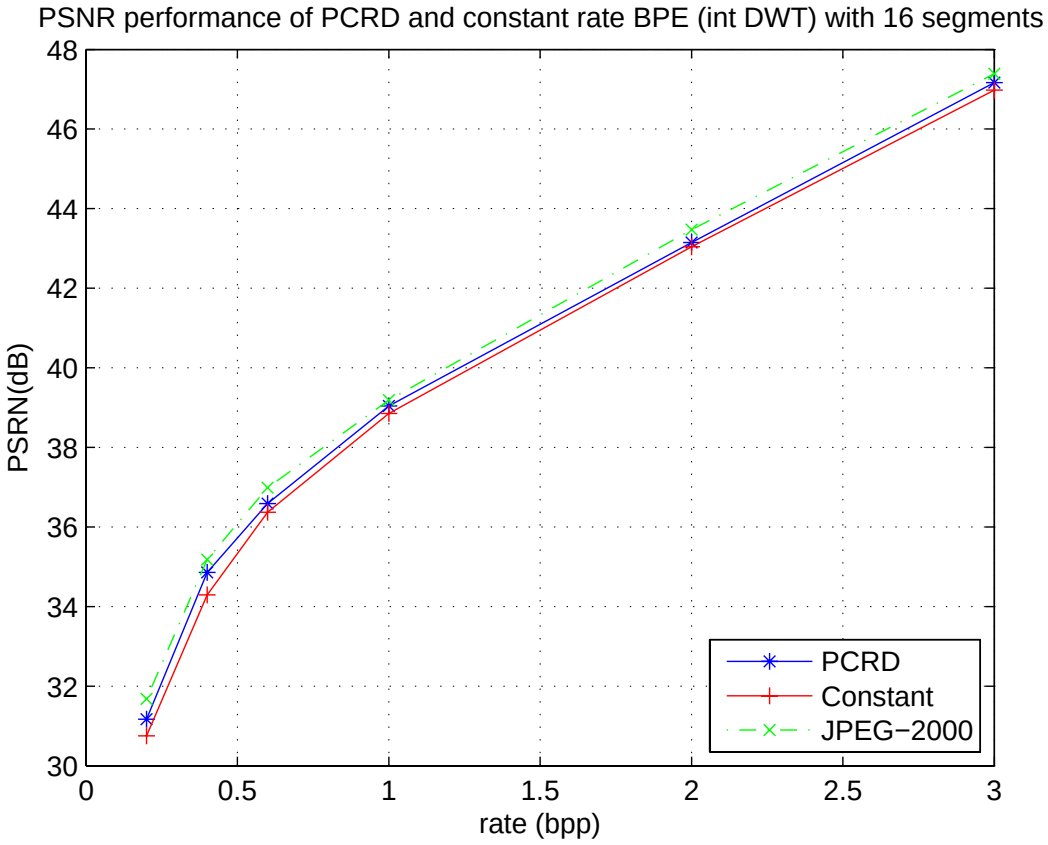


Figure 3.9: PSNR (in dB) of CR-BPE, PCRD-Opt BPE, and JPEG2000, where the number of segments is 16

The modified rate control model using joint ρ domain and PCRD-Opt algorithm is applied to the BPE coder. The regular BPE is performed on the first three bit planes. Once the rates of highest three bit planes are available, they are used to derive the rate control model for the lower bit planes. The pre-determined distortion reduction, the PCRD-Opt algorithm is applied to the bits planes that have not yet actually been coded.

Rate	0.2	0.4	0.6	1.0	2.0	3.0
CR-BPE	30.55	34.54	36.75	39.44	44.00	47.93
PCRD-BPE	31.28	35.17	37.12	39.79	44.25	48.10
ρ-BPE	31.22	35.03	36.97	39.60	44.17	47.99

Table 3.2: PSNR (in dB) comparison using the ρ-PCRD-Opt algorithm (floating point DWT)

Rate	0.2	0.4	0.6	1.0	2.0	3.0
CR BPE	30.82	34.35	36.42	38.92	43.09	46.99
PCRD-Opt BPE	31.20	34.99	36.59	39.16	43.17	47.21
ρ-PCRD-Opt BPE	31.17	34.93	36.52	39.06	43.11	47.12

Table 3.3: PSNR (in dB) comparison using the ρ-PCRD-Opt algorithm (integer DWT)

Table 3.2 and Table 3.3 show the testing results for floating point DWT based BPE and integer DWT based BPE, respectively. The results indicate that for both the floating point DWT based BPE coder and the integer DWT based BPE coder, there is coding gain over the CR-BPE. In addition, the results show that when the bit rates are low, the PSNR performance is closer to the original PCRD, because the highest three

bit planes and bit planes with more accurate prediction are used for rate allocation. When the rates become higher, the predicted rates start working for the PCRD-Opt algorithm very well. Under this scenario, the savings over the original PCRD-Opt algorithm is that we do not have to complete the encoding of all bit planes.

3.4 Conclusions

In this chapter, a new rate allocation method for wavelet transform image coders is proposed. The PCRD-Opt algorithm requires the complete coding of the image in order to construct the rate distortion curve. It is a computationally demanding process suitable for offline coding applications. For applications subject to constraints of complexity and computational resources, this needs to be simplified.

The ρ domain analysis has been examined, and then adapted to facilitate the PCRD-Opt algorithm. ρ domain analysis does not directly build the model between rate and quantization, but rather a linear model between rate and percentage of zeros ρ in each bit plane. This linear model turns out to be accurate at higher order bit planes, i.e., when the quantization parameters are large. An improved model is developed for lower order bit planes, or equivalently small quantization parame-

ters. Incorporated with this ρ domain model, the PCRD-Opt algorithm is employed to find the optimum rate allocation. The BPE coder has been used to test the proposed algorithm. The PSNR performance is encouraging when compared with the CR-BPE, and approaches to the performance of the PCRD-Opt algorithm significantly reduced coding complexity and computational costs.

Chapter 4

Region Based Image Coding with Rate Control

This chapter focuses on region based wavelet image coding. If an image can be segmented to several regions that contain different level of activity, details, and texture information, the region that contributes more to visual quality can be compressed using more coding resources and transmitted with a higher priority than the other regions. This is referred to as region-of-interest (ROI) processing. The ROI processing allows more flexibility on manipulation of image content and transmission, and therefore, it provides a higher premium on intelligence, flexibility, robustness, and efficiency than the traditional coding.

In this chapter, we propose a simple yet efficient method to find the

ROI in the DWT transform domain. The BPE coder is extended to accommodate the concept of ROI with adaptive rate control. The goal is to assign more bit resources to the ROI using the proposed rate allocation method such that robustness can be enhanced without incurring the penalty of rate-distortion performance. Syntax modification to the BPE coder is proposed in order to use the ROI and the modification will be used in the next chapters.

The chapter is organized as follows: in Section 4.1, the concept of region-of-interest (ROI) is discussed; in Section 4.2 a segmentation algorithm in the DWT transform domain is proposed; in Section 4.3, the proposed algorithm is applied to the BPE coder, and syntax modification of the BPE coder to incorporate ROI is recommended; in Section 4.4 rate control schemes proposed in Chapter 3 are applied and the performance is compared with CR-BPE coder and the regular BPE coder; a summary of the chapter is presented in Section 4.5.

4.1 Region-Of-Interest (ROI)

In image and video compression applications, many sophisticated methods are developed to describe the ROI, such as the shape coding and object based coding in MPEG-4. To find and describe the ROI using those

techniques can be computationally expensive, as mathematical models may be required in order to accurately describe a variety of irregular shapes. Furthermore, rapid scene changes in video sequences mean that these models would have to be employed repeatedly thus increasing the computational burden. These factors restrict these methods from wide deployment for real time applications which operate under complexity constraints.

In this chapter, we do not intend to develop complex mathematical models to describe objects for compression. Rather, we intend to develop a simple and efficient method to find the ROI of an image in the wavelet transform domain, incorporated with the bit plane coding of the wavelet image coder. Here the ROI may not have the same interpretation as the traditional ROI defined in MPEG-4 or JPEG2000. In this application, the ROI serves not only for the potential robust transmission, but also for the rate control in the wavelet transform domain. The rate allocation is performed in this scenarios such that more bits are allocated to the ROI that has more details and activities in the transform domain.

4.1.1 Region-of-Interest (ROI) and Rate Control

It is reasonable to assume that non-homogenous regions are more difficult to compress than homogenous regions. For instance, in the BPE coder we have to use more bits to differentially encode the DC coefficients and AC depth. In addition, for encoding of AC coefficients, if very few AC coefficient stand out from the other coefficients, many extra bits are needed to indicate the significance of AC coefficients in those blocks with small AC coefficients.

Intuitively if an image contains more homogenous regions, rate allocation schemes may work better than an image with less homogenous regions. A ROI can be assigned more coding resources than the non-ROI. Even if the non-ROI is discarded, there may be no severe visual degradation within the ROI provided that the ROI that contains the desired detail is available.

4.1.2 Region-Of-Interest and Robustness

From the perspective of robustness, if we know the ROI, then many error protection schemes can be applied directly to the bits representing the ROI and protect them against bit errors. Forward error correcting (FEC) codes such as Turbo codes, LDPC codes can be employed for the

ROI. This approach is usually referred to as unequal error protection (UEP).

4.1.3 Partial and Progressive Decoding

Once the ROI is located, it can be encoded and then transmitted with higher priority. Once bits for the ROI have been received, the transmission can be terminated and the bit stream can be truncated if necessary in order to save bandwidth and coding resources.

4.2 Segmentation of ROI

The ROI may be determined by end users based on their subjective observations. Alternatively, from the perspective of compression and rate control, the ROI can be automatically determined based on some objective criteria. Therefore, we can simply categorize the process as follows.

- User Defined Region-Of-Interest (UDROI). The ROI is dependent on the end users. These regions may not necessarily be the regions that have the most activity or details.

- *A*utomated *S*egmentation *R*egion *O*f *I*nterest (ASROI). The ROI

can be located by analyzing the images using edge detection algorithms, and morphological algorithms.

In this section, an algorithm for ASROI is proposed. Note that the main goal is to facilitate compression and rate allocation. The regions that contain activity and details are supposed to take relatively more bits. While the regions that contain homogenous or consistent backgrounds are regarded as regions of non-interest (Non-ROI). Relatively fewer bits are allocated to the non-ROI and they also have a lower transmission priority compared with the bits from the ROI.

4.2.1 Algorithm

We use a region growing method to find the ROI. In particular, a number of 8×8 blocks with similar properties are grouped together as follows.

1. **Find a proper seed block:** a seed block is chosen by checking the transform coefficients. The block that has the maximum sum of the absolute values of AC coefficients is picked up as the seed block. The statistical properties of this block are taken as the initial properties of the region.

2. **Iterative region growing:** assume the coordinate of the current

block is (x, y), the following function recursively checks if the neighboring blocks are in the region.

```
RegionGrow(x, y, region)
{
If(isXYinRegion(x, y, region) == TRUE)
updateRegion(x, y, region);
endif
RegionGrow(x - 1, y, region);
RegionGrow(x, y-1, region);
RegionGrow(x + 1, y, region);
RegionGrow(x , y + 1, region);
}
```

where *isXYinRegion* is a function to check if the block is in the region. If it is in the region, the function *updateRegion* is used to update the region information. In *isXYinRegion* function, we determine if the block is in the region based on some thresholds. Fixed thresholds may result in the failure of properly assigning blocks to the ROI. To solve this problem, the thresholds are dynamically changed based on the region information. Then we perform a bidirectional search by updating the threshold and checking if there are changes in the size of the ROI. If the

change in the size of the ROI is relatively small, then we assume the algorithm has converged and the ROI has been found.

4.2.2 Grouping Criteria for Region Growing

To find the optimum segmentation in terms of coding rates may not be trivial. An exhaustive algorithm may search for every possible permutation of segmentation and then the optimum can be found. However, this is not feasible in this work and therefore is not in the scope of this dissertation.

To determine if a block belongs to the ROI, several properties and statistical parameters could be examined to see if they match each other. The mean average error ΔMAE is found to be accurate and to produce good grouping results, which is defined as the difference between the candidate block and the region as follows:

$$\Delta\text{MAE} = \sum_{i=1}^{8}\sum_{j=1}^{8} |p[i][j]| - |p_r[i][j]| \tag{4.1}$$

where $|x|$ represent the absolute value of x, $p[i][j]$ is the coefficient at $[i, j]$ in the candidate block, and $p_r[i][j]$ is the average of the pixels at

$[i, j]$ of all r blocks, i.e.,

$$p_r[i][j] = \frac{\sum_{k=1}^{r} |p_k[i][j]|}{r} \tag{4.2}$$

If ΔMAE is below a threshold, this block is grouped into the region. Otherwise, it is regarded as a block in region of non-interest.

4.3 BPE Coder with ROI (BPE-ROI)

It is not straightforward to use ROI in the BPE coder as there is no syntax in the BPE coder defined to handle different regions. Therefore, extra coding syntax and an ROI mechanism have to be used. In the BPE coder, the basic coding unit is a gaggle, which consists of up to 16 8×8 blocks, and a number of gaggles are grouped to produce a segment. In the BPE coder, blocks in gaggles and gaggles in segments have to be consecutive. However, shape of the ROI may be arbitrary and blocks in the ROI may not be consecutive. To accommodate these changes, the header structure of the BPE coder needs to be modified and masking bits are needed. The encoder needs to reorder the blocks such that the blocks in the ROI are encoded first. The decoder then reorganizes the decoded blocks to their original order for reconstruction. These modifications,

however, do not affect the main coding structure of the BPE coder. We call the BPE using ROI the BPE-ROI.

4.3.1 Significant Block Map (SBM)

A **Significance Block Map** (SBM) is used as side information to indicate whether blocks are in the ROI or not. One masking bit is sent out for each block. The bits in SBM are sent in raster order and immediately after part 4 of the header of the first segment and before coding of the blocks. Given an image of size 512×512, the number of 8×8 blocks is $512 \times 512/(8 \times 8) = 4096$. Therefore the cost introduced is $\frac{1}{64} = 0.015625$ bits/pixel if the SBM is sent out without entropy coding.

In natural images the homogenous regions are more likely connected, and we expect that 1s and 0s in SBM occur consecutively, and therefore, the overhead can be lowered with run-length coding or other advanced methods of entropy coding. As coding bit rates increase, the cost of coding those overhead bits becomes relatively low and the coding penalty is negligible.

0	0	0	0	0
0	0	1	1	0
0	0	1	1	0
0	0	1	1	0
0	0	0	0	0
0	0	0	0	0

Figure 4.1: ROI

7	8	9	10	11
12	13	1	2	14
15	16	3	4	17
18	19	5	6	20
21	22	23	24	25
26	27	28	29	30

Figure 4.2: SBM of the ROI

4.3.2 Modification of Header

Two new coding modes are specified by modifying the segmentation header. The reserved bits available in the BPE syntax can be used to carry extra information to the decoder. As shown in Table 4.1, a new part 5 in the header is defined. Figure 4.3 shows how the new defined header and SBM work with the original BPE.

Header	parameters	original	proposed	representation
1	P1_Reserved_1Bit	1	1	0: regular coding, 1: region coding
2	P2_Reserved_4Bits	4	2	0: region grow, 1: rectangular region
5	BLOCKS_HEIGHT	12	12	Block Height
5	ROIWeight	3	3	Bits shift up
5	MotionEstimation	1	1	Reserved for video coding

Table 4.1: The modified header in the BPE coder for coding with ROI

In part 1 of the header, P1_Reserved_1Bit is redefined as a switch.

If it is set to 1, the region based coding is initiated. Otherwise it is in regular coding mode.

If P1_Reserved_1Bit = 1, P2_Reserved_4Bits is used for switch between the ASROI and the UDROI. If it is a rectangular UDROI, the header 5 records its height. ROIWeight in header 5 is an optional weighting factor for the ROI. Similar to JPEG2000, the pixels in the ROI can be shifted up by ROIWeight number of bits such that the ROI can be encoded with higher priority for bit scanning.

4.4 Experiment and Results

In this section, we first demonstrate the region grow method and the dilation operation to connect the separated blocks into a region. Two rate allocation methods, the PCRD-Opt algorithm and the ρ-PCRD-Opt algorithm, are applied to the BPE-ROI. We demonstrate the effectiveness of the region segmentation algorithm in rate allocation applications.

4.4.1 Demonstrations

Figure 4.5 demonstrates the results of using the region growing method to segment the original Lenna image as shown in Figure 4.4, where the white region represents the ROI region found using the algorithm and the

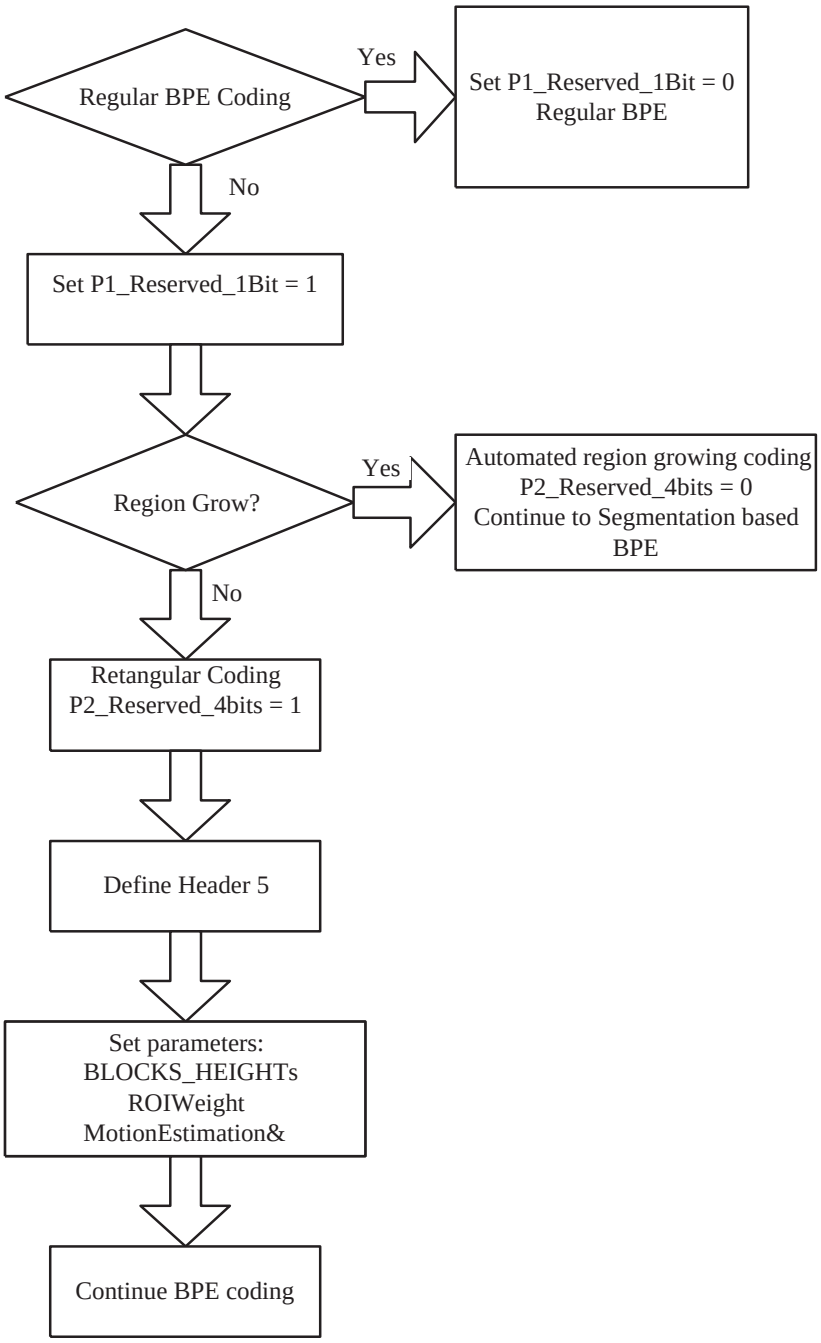


Figure 4.3: Flowchart defining bits and header for the BPE-ROI

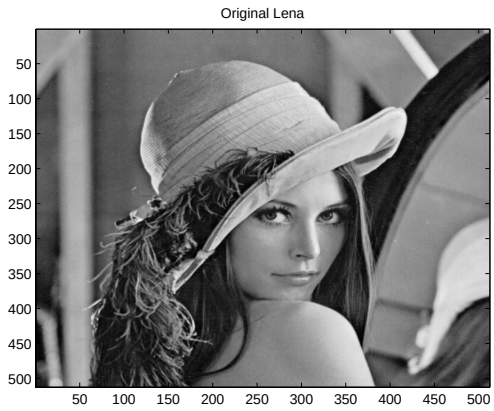


Figure 4.4: The original Lenna image

black region represent the non-ROI region. We can see that the blocks containing more activity and details are grouped into one big region. The rest of the blocks are mostly blank and have little detail.

Figure 4.5: The segmented Lenna in the DWT domain

We can see that there are several blocks isolated by the blocks in the ROI. Though these blocks are less active than the blocks in the ROI, their absence or lower rate allocation could have a negative impact on

the visual effect of the ROI. To mitigate the effect, a dilate operation can be employed to remove or reduce the number of isolated blocks. The dilate operation is a common operation which is generally used for binary images to enlarge the boundaries of regions and reduce holes in regions. This is equivalent to a low-pass filtering operation and its effect can be seen in Figure 4.6.

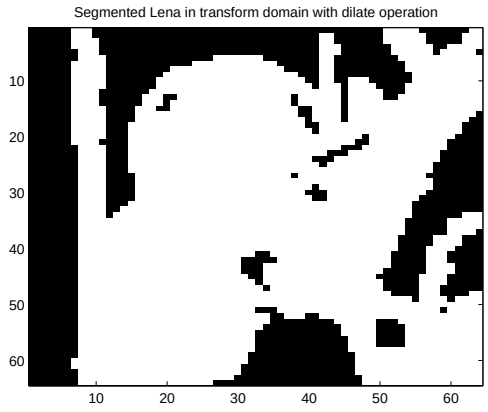


Figure 4.6: The segmented Lenna in the DWT domain with dilate operation

4.4.2 Constant Rate BPE Coder (CR-BPE) with ROI

The compression performance of the BPE-ROI is compared with the regular BPE coding without region classification and we want to see if there is any PSNR improvement given the same bit rates. First the ROI and the non-ROI are extracted using the iterative region growing algorithm. For both regions, every 256 blocks are grouped as one segment.

If the number of blocks in ROI and non-ROI are not the integer multiples of 256, we do rounding operation such that the remainder blocks create an independent segment if it is greater than 128, otherwise they are attached to the last segment. To make the comparison fair, for tests with the regular CR-BPE the 512×512 Lenna image is divided into 16 segments and each segment contains 256 blocks in raster scan order.

Initially we expected that the constant rate BPE-ROI could yield better compression performance than the regular constant rate BPE. However, it turns out that the PSNR performance is far worse (1.2dB) than the CR-BPE, as shown in Table 4.2.

Rate (bpp)	ROI	regular
0.2	29.69	30.55
0.4	33.40	34.54
0.6	35.36	36.75
1.0	38.00	39.44
2.0	42.60	44.00
3.0	46.96	47.93

Table 4.2: PSNR (dB) performance of the regular CR-BPE and the CR-BPE with the ROI (Floating point DWT, 16 segments)

By analyzing the regions, it is easy to see why this happens. The ROI contains much more activity than the non-ROI. The CR-BPE allocates bit resource into each region regardless of their activity. In this scenario, given a bit plane, the bits in the ROI are not coded, while the same bit planes in non-ROI may be coded. Therefore, more distortion

is contributed by the ROI regions than the non-ROI regions, and the overall distortion is therefore increased.

From another point of view, the drop of PSNR actually in turn implies that the iterative region growing algorithm works as expected and the ROI we have found makes sense. Otherwise, the PSNR would not have such a big drop.

4.4.3 Rate Control for the BPE Coder with ROI

Since CR-BPE does not produce coding gain, it is necessary to adaptively allocate bit resources. The rate allocation methods developed in Chapter 3, the PCRD-Opt BPE coder and the ρ-PCRD-Opt BPE coder, can be directly applied. For each coding segment, we build the distortion curve over rate reduction and then find the optimum truncation points. The experimental results are presented as in Table 4.3:

Rate	CR ROI	CR NR	PCRD NR	ρ-PCRD NR	PCRD ROI	ρ-PCRD ROI
0.2	29.69	30.55	31.28	31.22	31.14	31.01
0.4	33.40	34.54	35.17	35.03	35.19	35.07
0.6	35.36	36.75	37.12	36.97	37.23	37.05
1.0	38.00	39.44	39.79	39.60	39.88	39.67
2.0	42.60	44.00	44.25	44.17	44.32	44.18

Table 4.3: PSNR (dB) performance comparison floating point DWT, 16 segments: where CR-ROI represents constant rate ROI without adaptive rate allocation, CR NR represents constant rate without ROI. PCRD NR represents PCRD algorithm without ROI, ρ-PCRD NR represents PCRD algorithm without ROI, and PCRD ROI represents the PCRD with ROI, ρ-PCRD ROI represents the ρ-PCRD with ROI,

As we can see, the PSNR performance of the PCRD-Opt BPE-ROI and the ρ-PCRD-Opt BPE-ROI has been improved over the BPE without ROI at middle to high rates by around 0.1dB. This is not as much as for low bit rates. Actually, at low bit rates, the performance is actually lowered. This is, because the coding of SBM takes a portion of bit resource. When the bit rate is low, this overhead can compromise the coding gain. Another explanation is that at high bit planes, the symbols obtained from the bit scanning of AC coefficients in the ROI tend to be zeros, and therefore, the symbol pattern similarity among blocks is not good enough to produce coding gain using the entropy coding.

4.4.4 Discussion

As we have discussed, the goal of the BPE with ROI are not solely for compression, but also for better control of the content and visual effect. As far as the PSNR performance is concerned, the PCRD-Opt BPE-ROI and the ρ-PCRD-Opt BPE-ROI have demonstrated competitive performance as compared with the regular PCRD-Opt algorithm and the ρ-PCRD-Opt BPE without ROI. It is reasonable to assume that the ROI has more impact on the visual effect, as seen from Figure 4.6, and in this scenario, assigning more coding resources to the ROI may not

necessarily minimize the overall distortion. The benefit, however, is that the ROI can be separated such that the image content can be better manipulated. In addition, as the ROI can be transmitted with higher priority, more flexility for flow control by end users is achieved.

4.5 Conclusions

In this chapter, we have proposed the ROI concept in the transform domain, and an iterative region growing method to find the ROI by examining the similarity of transform coefficients in the DWT domain. Rather than classifying or describing objects using complex mathematical models, we classify regions based on the activity in the transform domain and then reorganize them into regions for efficient rate allocation.

The concept of ROI is applied to the BPE coder and the corresponding syntax modification is recommended. We apply the rate control methods proposed in last chapter to rate allocation using ROI. Competitive compression performance is demonstrated as compared with the regular BPE coder, while with benefits of more flexibility on ROI manipulation and transmission, and potential robustness if UEP is used. The CR-BPE with ROI is inferior than the regular CR-BPE without ROI in PSNR,

which in turn demonstrates the efficiency of the iterative region growing algorithm.

Chapter 5

Motion Image Coder Based Video Coding

Conventional video coders such as the MPEG-1,2, and 4 employ motion compensation to reduce redundancy and achieve excellent rate-distortion performance, while intra-frame video coding based on image codecs such as Motion JPEG/JPEG2000 has been widely adopted in many applications due to various reasons. In this chapter, we are interested in extending the wavelet based image coder for intra-frame based video coding without using motion compensation, and it serves as a basis for the robust video coding scheme proposed in Chapter 6.

This chapter is organized as follows: intra-frame and inter-frame video coding are discussed in 5.1; rate control methods are proposed for this

application based on the ones developed in Chapter 3 and applied to the BPE coder for intra-frame video coding in Section 5.2. The conclusion is presented in Section 5.3.

5.1 Inter-frame and Intra-frame Coding

The conventional inter-frame motion-compensated video coders that employ motion compensation have a dominant position in video coding applications. However, many high-speed/high-definition video coding systems employ intra-frame video coding schemes that use existing image coders as basis for video compression, such as Motion JPEG/JPEG2000. Though its PSNR performance is generally inferior to that of inter-frame video coding, intra-frame video coding based on image coders has been widely accepted due to several factors as follows:

- Despite the fact that many optimized and fast algorithms have been used, motion compensation still requires a considerable amount of time and computational resources. In many applications, it can take up to ninety percent of the coding time (22). For high-speed/high-definition image and video applications, requirements on complexity, delay, and energy are relatively high, and motion estimation and

motion compensation may not be affordable. As a consequence, intra-frame video coding is more feasible than inter-frame video coding in these scenarios.

- Motion compensated inter-frame video coding may suffer from the drifting effect discussed in Section 2.1. Intra-frame video coding does not have this issue as each frame is coded independently. Therefore, it is more robust and can be used in applications where robustness is a major concern.

- The cost of coding motion vectors may compromise the gain of using motion compensation. For slow motion pictures, the motion vectors can be small and they can be coded effectively using predictive coding. However, the cost of coding motion vectors for heavy motion video sequences and high-definition video sequences can be high if the dynamic range of motion vectors is large. Mode decision usually has to be developed to determine whether inter or intra mode should be used, which generally requires multi-pass optimization and incurs extra computational load.

- Research shows that for high-definition video sequences, using intra-frame video coding such as the motion JPEG2000 can achieve com-

parable coding performance as those inter-frame video coders such as H.264 (22). One of the explanations is that high-definition pictures contains a lot of details and variations and the resulting residual after motion search and motion compensation may be hard to compress.

Motivated by these factors, we attempt to extend the wavelet based image coder for intra-frame video coding without using motion compensation. The methods and observations will be used in Chapter 6 for the development of robust video coding.

5.2 Motion Image Coding with Adaptive Rate Allocation

In motion image based video coding, no motion estimation or compensation is performed and each frame is coded independently. The problem we are interested in is how to properly allocate bit resources to each individual frame such that the total distortion is minimized for given consecutive frames.

This problem is similar to the rate allocation problem addressed in Section 3.1, where for each segment a number of bits are allocated to

minimize the total distortion for single images. The difference is that now we have two levels of coding units: frame level and segment level. If each frame is treated as a segment, then the rate control only involves frame level optimization. If each frame needs to be divided into small segments for compression, then segment level optimization is necessary.

Ideally both frame and segment levels are required to achieve an optimized performance if each frame needs to be divided into small coding units. We test the coding of frame level and segment level separately. In the experiments, we use the BPE coder as the wavelet image coder and the Crew sequence as the testing sequence. As shown in Figure 5.1, each frame in the sequence has a resolution of 1280×760. We test 50 frames (luminance component) and set the data rate to 0.4 bits/pixel.

5.2.1 Constant Frame Rate and Variable Segment Rate (CFR-VSR)

Constant bit rates are assigned to each frame, while each frame is divided to small coding units for rate optimization. The rate control algorithms developed in Chapter 3 are applied, and the PSNR gain over the CR-BPE is computed.

We divide each frame into 16 segments, i.e., each segment contains

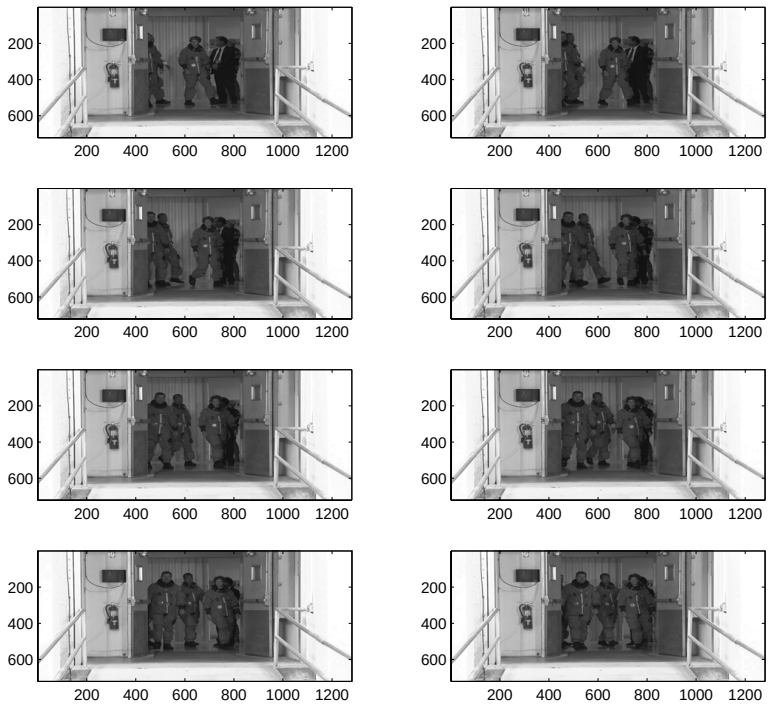


Figure 5.1: Samples of Crew sequence (one out of every five frames)

Schemes	PSNR improvement over CBR
VBR	+0.88
VBR-ρ	+0.61

Table 5.1: PSNR improvement of BPE using constant frame rate and variable segment rate versus the constant bit rate

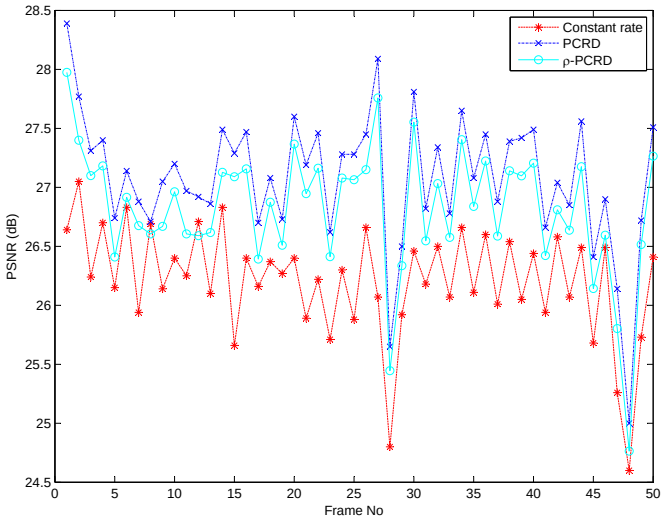


Figure 5.2: Rate control performance comparison using constant frame rate and variable segment rate)

$1280 \times 760/(64\times) = 950$ blocks. The PCRD-Opt BPE coder and the ρ-PCRD-Opt BPE coder are applied to both experiments. The PSNR performance is shown in Figure 5.2. As we can see, the PCRD-Opt BPE coder is much better than the CR-BPE coder, and the average PSNR gain for the 50 frames is 0.88dB. The ρ-PCRD-Opt BPE coder, though not as good as the PCRD-Opt BPE coder, also achieves better PSNR gain than the CR-BPE coder, and the coding gain is 0.61dB. Table 5.1 summarizes the results. The ρ-PCRD-Opt BPE coder does not require the computation of the complete R-D curve, and, therefore, has a much lower computational cost when compared with the PCRD-Opt BPE coder.

5.2.2 Variable Frame Rate and Single Segment Rate (VFR-SSR)

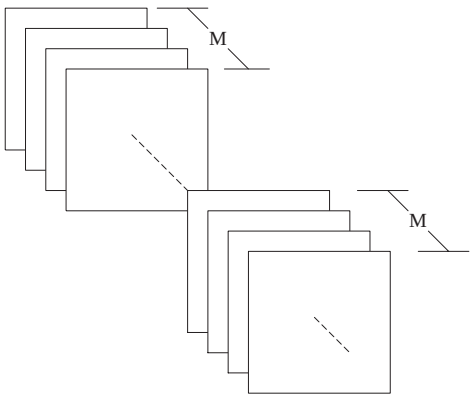


Figure 5.3: Frame level rate allocation

In this experiment, every block of M frames is organized into one coding group. Given the target bit rates for the group, the rate allocation is performed within the group, as shown in Figure 5.3. The coding rates are fixed for every M frames, while each frame may be allocated different number of bits, depending on the rate allocation schemes.

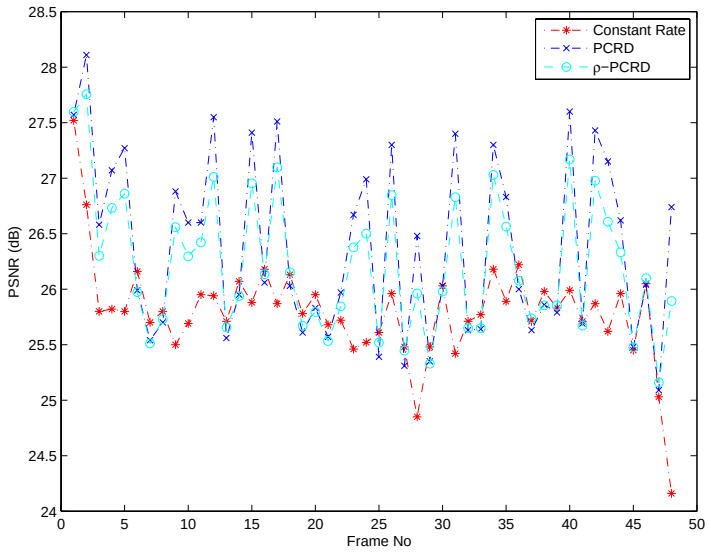


Figure 5.4: Rate control performance of Crew (Variable Frame Rate and Single Segment Rate)

Schemes	PSNR improvement over CR-BPE
VBR	+0.58
VBR-ρ	+0.41

Table 5.2: PSNR improvement of BPE using variable frame rate and single segment rate (VFR-SSR) VS CR-BPE

We set M to 4, and apply the PCRD-Opt algorithm and the ρ-PCRD-Opt algorithm to the frame level rate allocation. The PSNR performance

is shown in Figure 5.4. As we can see, the PCRD-Opt BPE coder is better than the CR-BPE, and the average PSNR gain for the 48 frames is 0.58dB. The ρ-PCRD-Opt BPE coder also achieves a PSNR gain of 0.41dB over the CR-BPE coder. Table 5.2 summarizes the results. While there is an average coding gain for the ρ-PCRD-Opt BPE coder, for certain frames, there is no coding gain. Since the rate allocation is performed among different frames, therefore, for certain frames, the rate allocated may be less than the rate allocated from the CR-BPE coder.

5.2.3 Variable Frame Rate and Variable Segment Rate (VFR-VSR)

We assume that each individual frame is divided into a fixed number of coding units. The objective is to allocate bit resources for the M-frame groups among the segments, such that the overall distortion is minimized. In this case, frame- and segment-level rate allocations are needed. These two allocations may be jointly performed. We can simply take all segments from each frame and group them together. The rate allocation can be obtained by applying the PCRD-Opt algorithm and the ρ-PCRD-Opt algorithm to this entire group. We call this joint frame and segment rate allocation (JFS-RC).

Alternatively, frame- and segment-level rate allocations can be separated into two stages. The first stage is to allocate bit resource among frames. The second stage is the rate allocation among segments. Here, we proposed a method to adaptively allocate rate based on the rate allocation of the past frames.

1. For the first frame in the M frame group, a full encoding is done. Its rate reduction, Δr, and distortion reduction, Δd, are recorded. Then a ρ curve is obtained.

2. For the remaining $M-1$ frames in the group, the rate can be estimated by using the ρ curve obtained in step 1. Assume the curve obtained is as follows,

$$R(\rho) = \alpha(1-\rho) + \beta \tag{5.1}$$

   The estimated rate is given by

$$R(\rho') = \alpha(1-\rho') + \beta \tag{5.2}$$

   where the ρ' is the ratio of the number of coefficients that are zero to the total number of coefficients.

3. Using the method in Subsection 3.2, the frame level rate allocation

is completed.

Note that here, we assume that the ρ curve is fixed for the M frames. Therefore, α and β are fixed for the M frames. For the frames in the M frame group except the 1st frame, the bit rate is obtained by calculating ρ' at at each bit plane, and then evaluating R using Equation 5.2.

After the bits are allocated at the frame level, we proceed to the second stage which is the segment level rate allocation, which can be done either using the PCRD-Opt algorithm or the ρ-PCRD-Opt algorithm. Here for comparison purpose, we applied the PCRD-Opt algorithm for JFS-RC. This yields an improvement in the PSNR, because all the segments are combined and the rate is jointly optimized.

In addition, we also fixed the rate for each frame to 0.4 bits/pixel and then applied the PCRD-Opt algorithm for each rate. The average coding gain in terms of PSNR using joint frame and segment rate allocation (JFS-RC), 2-stage PCRD, 2-stage PCRD-ρ over single frame using the PCRD-Opt algorithm is 0.94dB, 0.77dB and 0.50dB, respectively, as summarized in Table 5.3.

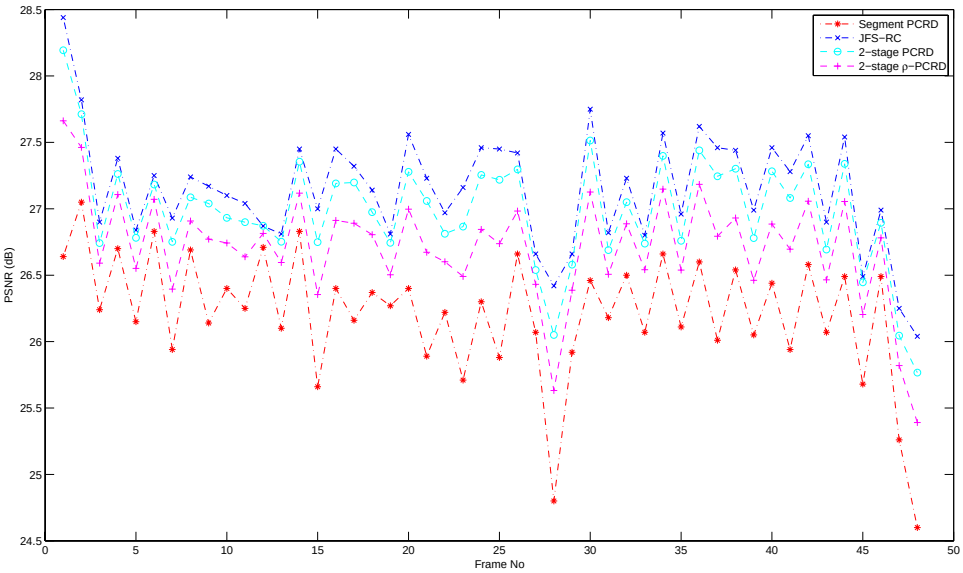


Figure 5.5: PSNR performance using 2-stage rate allocation (Variable Frame Rate and Variable Segment Rate)

Schemes	PSNR improvement over CR-BPE
JFS-RC	+0.94
2-stage PCRD	+0.77
2-stage PCRD-ρ	+0.50

Table 5.3: PSNR gain using joint frame and segment rate allocation (JFS-RC), 2-stage PCRD, 2-stage PCRD-ρ over single frame using the PCRD-Opt algorithm

5.3 Conclusion

In this chapter, the wavelet based image coder is extended for intra-frame video coding. The motivation is that intra-frame video coding has advantages over inter-frame video coding, such as robustness, speed, and complexity. In particular for high definition pictures, intra-frame video coding can save considerable computational load at small expenses of compression performance.

We examine and test rate allocation in different situations. The rate for each frame is fixed and then the rate allocation methods developed in Chapter 3 are applied to the segments in each frame (CFR-VSR). In addition, a fixed number of frames are grouped together and each frame is a coding unit. The rate is fixed for the frame group and each frame is set as a basic coding unit (VFR-SSR). The rate control schemes are applied to frame-level rate allocation. Lastly, frame and segment level rate allocations are performed jointly (VFR-VSR). A fixed number of frames are grouped and each frame is divided into small segments. The first rate allocation method is to group all segments and the optimization procedure is applied to these segments equally. A 2-stage rate allocation is to first allocate the rate in frame level by using the ρ-PCRD algorithm. The segment level rate allocation is then performed by using either the

PCRD-Opt algorithm and ρ-PCRD-Opt algorithm, respectively.

The wavelet based BPE coder is used for the testing. The results show a significant PSNR improvement. In the VFR-VSR we find that the PSNR can be improved over 0.94dB using the optimized rate allocation method, and the ρ-PCRD-Opt algorithm can achieve 0.50dB gain with much less computational cost.

Chapter 6

Robust Distributed Video Source Coding

In this chapter, video coding based on the wavelet image coder is developed by exploring the inter-frame redundancy without explicitly referencing any of the individual coefficients. With the knowledge of correlation models between consecutive frames, the concept of distributed source coding is applied for robust and efficient source coding.

The concept of distributed video coding is introduced in Section 6.1. A distributed video coding in wavelet transform domain is proposed in Section 6.2. It is then applied to the BPE coder in Section 6.3, the results are reported in Section 6.4. Section 6.5 concludes this section.

6.1 Distributed Video Coding (DVC)

As discussed in Section 2.3, based on the Slepian-Wolf coding (28) and Wyner-Ziv coding (29) theory, two correlated sequences can be encoded separately without sacrificing the coding efficiency as compared with joint encoding. This is often referred to as the distributed source coding (DSC).

As we know, generally neighboring frames in video sequences are highly correlated in the temporal domain. Given a frame to be encoded, we assume that side information can be inferred based on history frames. The pixels in side information are highly correlated to the pixels in the frame to be encoded, though their exact values in side information are not necessary known at the encoder. If an accurate correlation model between side information and the frame to be encoded can be found, video sequences can be effectively compressed without knowing the exact value of side information at the encoder using the principal of distributed source coding. In practice, the correlation models may be derived or estimated either on real time or by offline training. At the decoder, side information is derived by interpolation, extrapolation, or direct use of the past frames, and then is used for joint decoding to re-

construct the symbols that are encoded using distributed source coding.

The framework of distributed video coding is dramatically different from the traditional motion-compensated predictive video coders. Basically, there are two major applications using distributed video coding: one is for low-complexity video coding and the other is for robust video coding.

6.1.1 Low-Complexity Video Coding

In traditional video coders such as H.263 and MPEG-2, the encoders require 5-10 times more complexity than the decoders. Those video coders may not be suitable for applications where complexity and computational resources at the encoder are constrained, such as in a wireless sensor network. In this scenario, low complexity video encoders are desirable, even though the compression performance may be compromised.

In the past few years, many low complexity distributed video coding systems have been developed in both transform and pixel domains. Figure 6.1 shows a pixel-domain system proposed by Aaron et al. in (121; 122; 123). In this scheme, a frame is intra-coded using the conventional block based DCT encoding and intra-frame decoded without using any reference frames. We call this type of frame a key frame. Another

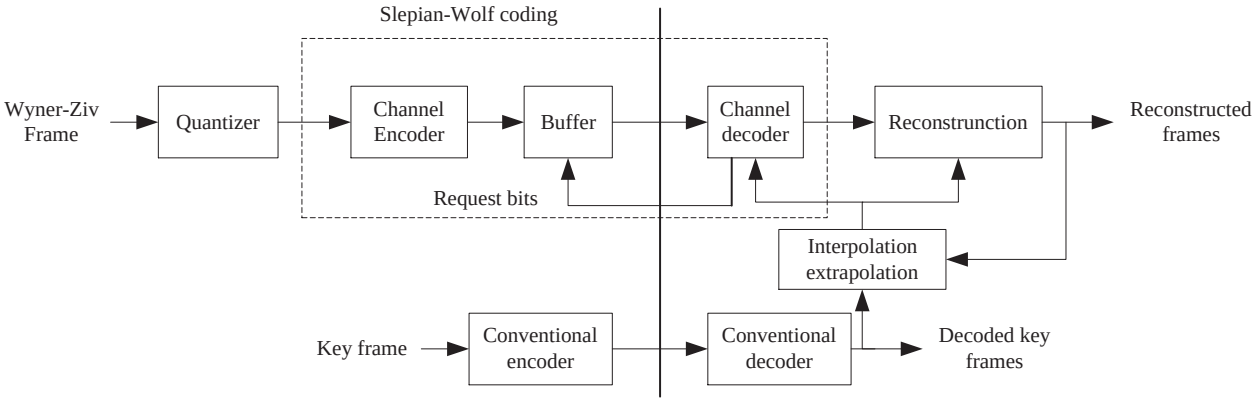


Figure 6.1: Low-complexity distributed video coding.

type of frame, the Wyner-Ziv frame, is intra-frame coded using a rate compatible punctured turbo code (RCPT) and inter-frame decoded at the receiver. Given every fixed number of input bits, the RCPT generates systematic bits and syndrome bits from the interleaved systematic convolutional codes and only sends the syndrome bits to the decoder, as illustrated in Figure 6.2 (1).

The decoder exploits the statistical dependency between frames by inter-frame processing to predict the pixels in Wyner-Ziv frame. More specifically, a correlation model is assumed, that is, the difference between the individual pixel values in Wyner-Ziv frame and the side information is assumed to follow Laplacian distribution. The Laplacian parameter is updated based on statistics obtained from previously decoded frames. A BCJR decoding algorithm, an algorithm for maximum

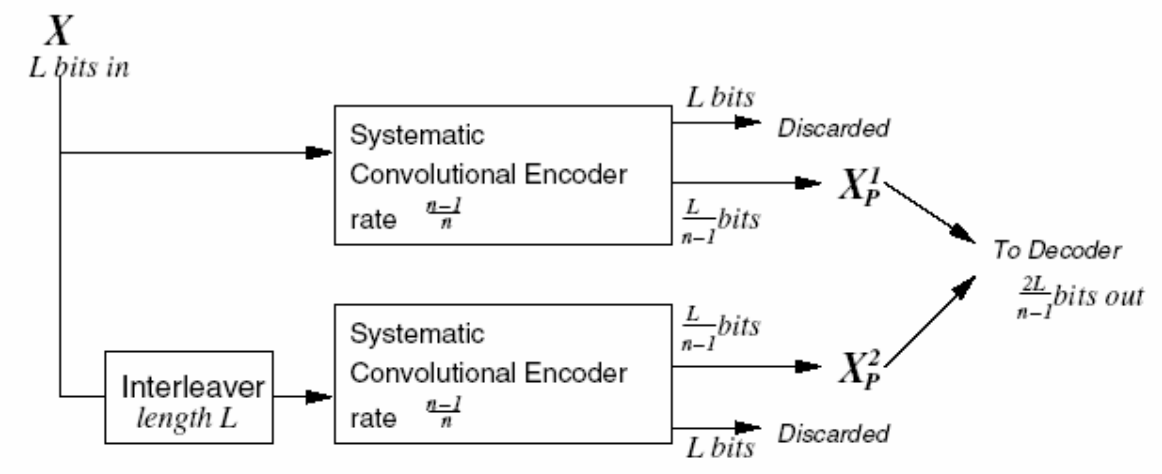


Figure 6.2: Codeword generation using Turbo-Code (1)

a posteriori decoding of error correcting codes, is employed to reconstruct the pixel in a Wyner-Ziv frame incorporating the correlation model. A feedback channel is needed for the decoder to request more bits from the RCPT code at the encoder if the decoder cannot reliably reconstruct the pixels based on the available number of bits. In this scheme, no motion estimation and compensation is performed. The major of the computational complexity is shifted from the encoder to the decoder as the BCJR algorithm at the decoder is more more complex than the operations at the encoder. It was shown that the compression efficiency is 2-5dB better than the intra-frame video coding, while there still a significant gap with traditional motion compensated video coders.

Distributed video coding has been extended to the transform domain

as well. A scheme called PRISM was proposed by Puri et al. (124; 125). In PRISM, a blockwise DCT is applied first to 8×8 blocks, followed by a uniform scalar quantization. Each block is then encoded independently. The least significant bit (LSB) portion of quantization indices that cannot be inferred from the side information is compressed using syndrome based BCH code, where the code parameters are derived based on correlation models from offline training. The most significant bit (MSB) portion of quantization indices are entropy coded. In this scheme, there is a simple motion estimation process, and the encoder sends a cyclic redundancy check (CRC) of the quantized coefficients to assist the motion compensation at the decoder.

Similar distributed video coding systems in the discrete wavelet transform (DWT) domain have been proposed in (126; 127), where the transform coefficients are coded by channel codes such as RCPT and LDPC codes.

Though these systems have reported excellent coding performance as compared with intra-frame video coding, there are several drawbacks:

- They generally require a feedback channel to communicate between the decoder and the encoder. In case the decoder cannot reliably infer the symbols to be decoded based on the side information and

available parity check bits, a request for more bits needs to be sent to the encoder. This process ends when the symbols are inferred with certain level of reliability. The feedback channel, however, is not always available in practical coding systems, and even if it is available, severe delay may be incurred which would make the system unusable for real-time video communication.

- Another issue is that the correlation models in these systems may require offline training. Those correlation models from training on one video sequences may not be accurate for the other video sequences, and for many real-time video coding applications, offline training may not be feasible as well.

6.1.2 Robust Video Coding

It is well-known that for conventional inter-frame video coders bit errors in a frame can propagate to subsequent frames that use pixels in this frame as references. This drifting effect can be catastrophic and severely degrade the decoded video quality. This can be limited by using distributed source coding techniques. The parity information can be used to enhance the error resilience performance. One method is to employ more powerful Slepian-Wolf codes, which not only detect the "error"

of the correlation channel between source sequence and side information, but also combat the errors in the bitstream. PRISM employs this method and the visual quality with loss of frames is better than that of H.263+. In (128), a redundant Wyner-Ziv frame is introduced in addition to the traditional frames to limit the error propagation. This is basically a joint source-channel coding technique, and it is reported that the robustness of the video coders was improved significantly.

6.2 Distributed Video Coding (DVC) in the Wavelet Domain

In this section we propose a robust distributed video system in the wavelet domain. Figures 6.3 and 6.4 show the encoder and the decoder of the system, respectively. The system is basically a hybrid coding system that combines the efficiency of distributed video coding and the robustness of intra-frame video coding.

The basic idea behind this hybrid coding system is that each coding block is classified into two categories: correlated or uncorrelated block, and the classification is determined by the correlation between the block and its co-located block in the reference frame. For uncorrelated blocks, intra-frame video coding is used, while correlated blocks are coded based

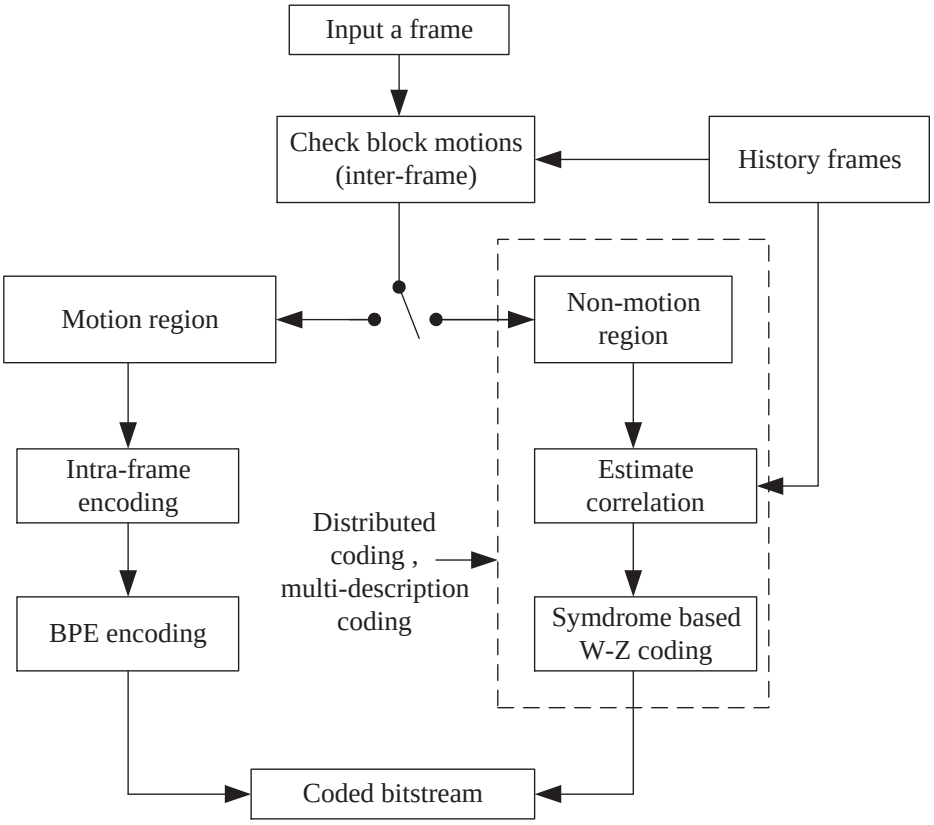


Figure 6.3: Encoder of robust distributed video system

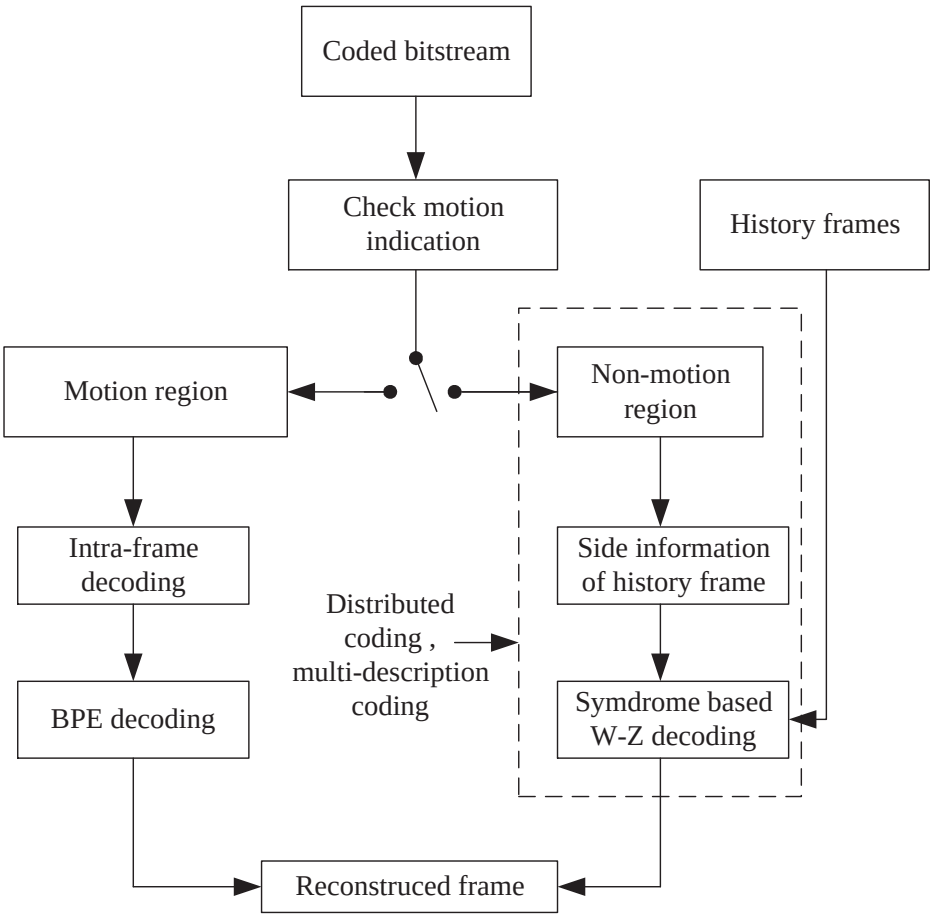


Figure 6.4: Decoder of robust distributed video system

on the principle of distributed source coding. We refer to uncorrelated blocks as motion blocks. However, strictly speaking, the occurrence of uncorrelated blocks may not be directly caused by motion, and some new scene or objects in a frame can cause associated blocks to be classified as uncorrelated blocks.

The approach avoids high complexity motion detection by either the encoder or the decoder and assumes that the motion blocks are more important for visual quality. The potential error propagation caused by erroneous bitstream can be also mitigated. Without motion estimation and compensation, we will be able to concentrate on building accurate correlation models for the non-motion blocks and using proper error-correcting code to improve the coding efficiency. The *Significant Bit Map* (SBM) described in Chapter 4 can be used to indicate whether the associated coding blocks are motion or still blocks. As compared with motion compensated video coding based on distributed source coding, we expect a sacrifice in coding efficiency. However, this compromise is reasonable as our goal is not only coding efficiency, but also robustness.

6.2.1 Block Classification and Correlation Model

To determine the correlation model for the blocks to be coded, some coding resources are allocated to compare the two co-located blocks in two neighboring frames. The motivation behind this comparison is that generally if in a video sequence the frame rate is sufficiently high and the scene changes slowly, two co-located blocks may be highly correlated. To examine the correlation, we may assume that there is a simple correlation model existing between the two co-located blocks as follows:

$$\hat{y} < \hat{x} + N \tag{6.1}$$

where $\hat{x}$ and $\hat{y}$ are the quantization indices of the co-located coefficients x and y, and N can be viewed as a correlation noise. We may assume that $|N|$ satisfies the following condition:

$$-\theta \leq N \leq \theta - 1 \tag{6.2}$$

θ indicates the correlation degree. A small θ means that the two blocks are more correlated, than two blocks associated with a large θ. Due to the variational nature of video sequences and different quantization steps, θ may vary over a wide range of values. If θ is large, then the cost of distributed source coding will be high, which violates the goal of

improving the coding efficiency. Therefore, we use θ as a criterion for determination of whether a block is a motion block or a still block. If θ is greater than a threshold θ_T, the block is then classified as a motion block.

At the encoder, if θ is known, without knowing $\hat{y}$, we can simply take the value of $\hat{x}$ modulo $2 \times \theta$, i.e.,

$$\hat{x}_\theta = \hat{x}(\text{MOD})(2 \times \theta) \tag{6.3}$$

$\hat{x}_\theta$ is transmitted to the decoder. The modulus operation is equivalent to grouping all possible $\hat{x}$ values into $2 \times \theta$ number of sets. Set 0 contains all $\hat{x}$ for which $\hat{x}_\theta$ is 0, and Set 1 contains all $\hat{x}$ for which $\hat{x}_\theta$ is 1, and so on. This is basically a binning process, and the index of a bin represents all the $\hat{x}$ that generates the same modulo. At the decoder, once we know $\hat{x}_\theta$, the candidate $\hat{x}$ are known. $\hat{x}$ can be reconstructed by simply taking the one that is closest to $\hat{y}$ in the set.

For example, if a set of modulo intervals are defined as follows:

$$S = \{[-1, 0], [-2, 1], [-4, 3], [-8, 7], \cdots, [-2^n, 2^n - 1]\} \tag{6.4}$$

and the code length of corresponding modulo intervals are $\{1, 2, 3, 4, \cdots, n\}$. Then for interval of $[-4, 3]$ and $x = 20$, the resulting value of x modulo

8 is 4, which is coded to 100 in a three bit binary representation.

The blocks are then classified based on the correlation interval they fall into. This can be examined in real time between two co-located blocks. Blocks are grouped in raster-order and the correlation information is transmitted before coding of each block. The interval θ is transmitted to the decoder along with the packet header.

6.3 Distributed Video Coding in the BPE Coder (BPE-DVC)

The well-known wavelet based image coders such as JPEG2000, SPIHT, and the BPE coder are all based on progressive bit plane scanning. However, the hierarchical structure used by bit plane scanning process in these coders is different. For instance, in JPEG2000, each individual subband obtained from the wavelet decomposition is split into code blocks comprised of 64×64 coefficients. As a result, the block classification scheme proposed is not suitable for the structure because code blocks are defined within one subband and only represent a certain type of frequency. In the BPE coder, each 8×8 block after regrouping represents all frequency components of the same geographical region in the original image, and therefore, the block classification method we propose is rel-

atively easier to fit into the coding framework of the BPE coder. Figure 6.5 shows the diagram of the proposed coding system.

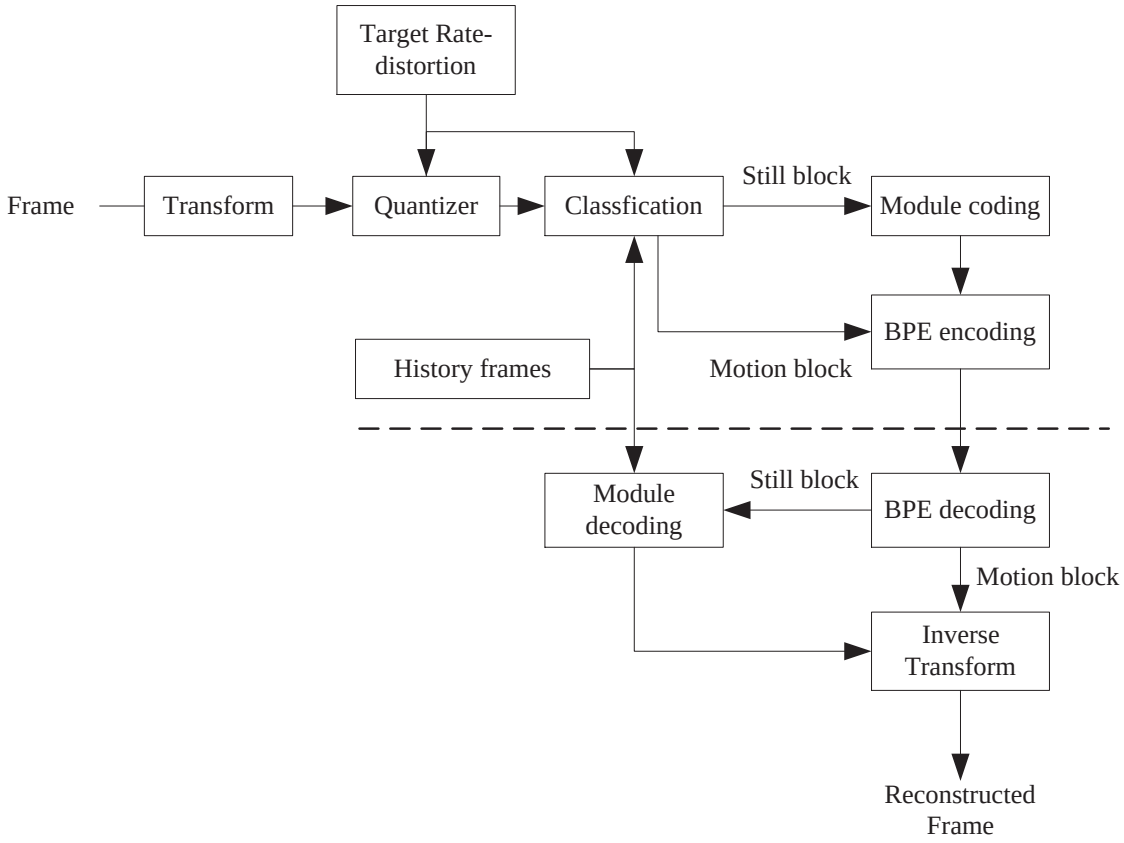


Figure 6.5: Proposed coding system based on the BPE coder

More specifically, the basic procedures of the encoding strategy are as follows:

1. DWT transform. We assume that there are enough memory buffers such that at least two frames are stored. After one frame is encoded, it is reconstructed in the transform domain and then stored

in the buffer. The next frame to be coded may use the previous reconstructed frame as reference.

2. Quantization and bit planes. The rate allocation solutions that have been developed in the last chapters can be used directly. In this scenario, we assume that not all frames are available. Given a target rate, we used the segment level rate allocation proposed in Chapter 5 to allocate the bit rate. The coding bit plane (i.e., quantization step) for the key frame is then determined and as this bit plane is set as the desired coding bit plane for the coding of the next Wyner-Ziv frame.

3. Correlation model and block classification. As described in Subsection 6.2.1, the blocks to be compressed in the current frame are compared with the co-located block in the reference frame to determine if the current block is coded as an intra block or an inter block that uses a correlation model. Given a block, the number of bins n is set to half of the difference of the highest and lowest bit planes. This process is continued until all the 8×8 blocks are classified. Note that the model is obtained based on AC coefficients.

4. After the block classification is completed, the significant bit map

(SBM) that is to indicate the type of the blocks is run-length coded and transmitted. In the SBM, each bit represents whether a corresponding block is intra coded or inter coded. If a type of blocks occurs consecutively, the SBM can be coded very efficiently.

5. Due to their importance to visual quality and vast dynamic range, DC coefficients and the AC depth are coded using the original differential coding techniques. Note that the AC depth here is not the original AC depth, but the AC depth after the modulo operation.

6. Coding of the model parameters. For each inter-frame coded block, its correlation parameter is coded similar to the coding of DC and AC depth coding of the BPE coder. The BPE coder is then applied to the coding of the bit planes.

7. Bit plane coding of the modulo symbols. The resulting symbols obtained from the modulo operation are coded by the BPE coder directly.

At the decoder, the decoding steps take similar operations of the encoding process in reverse order. In this system, at the encoder side, we employ the BPE coder, which is relatively more complex than the coding schemes used by other distributed coding systems. Meanwhile,

we employ a module based correlation model which is relatively more simple than the regular error control coding schemes incorporating the feedback channels in the distributed coding system. Hence, overall the proposed encoder may have a higher complexity than the encoders in the other coding systems. However, since at the decoder we do not have to perform exhaustive and iterative decoding which generally has to be done in other coding systems, the decoding complexity is much lower than the others. In addition, there are no feedback channels required. The rational behind this structure is to achieve a good balance between complexity and the performance, as well as a good complexity balance between the encoder and the decoder.

6.4 Experiments

In this section, simulation results that illustrate the performance of the proposed systems are presented. We examine the rate distortion performance of the BPE based coding system, and test the robustness of the system.

6.4.1 Compression Performance Experiment

First we illustrate the block classification process and intend to see how effective and accurate the coding block can be classified. In particu-

lar, we take 48 frames from the *Crew* and *shuttle* sequences to do the experiment.

Figure 6.6 and Figure 6.7 illustrate the two neighboring frames and the block classification results for *Crew* sequence, while Figure 6.9 and Figure 6.10 show the results for *Shuttle* sequence. As we can see in both sequences, the blocks in the current frame that are highly correlated with the co-located blocks in previous frame, especially the background regions, are classified as inter-blocks and they will be efficiently coded using the module operation based coding process. The rest of the regions, which contains a considerable amount of detail, will be directly intra-coded. As we can observe intuitively, the region coded by intra mode will have a significant impact on the reconstructed quality. In other words, a better protection of this region will benefit the reconstruction quality and robustness.

Figure 6.8 and Figure 6.11 show the ratio of the blocks that have been classified into intra-mode over the total number blocks in a frame. More specifically, from Figure 6.8 we can see that the percentage of the blocks classified into motion blocks in the second frame has been dramatically increased. By examining the frame, we observe that there is strong flashing effect presented in the frame, as shown in Figure 6.7. The

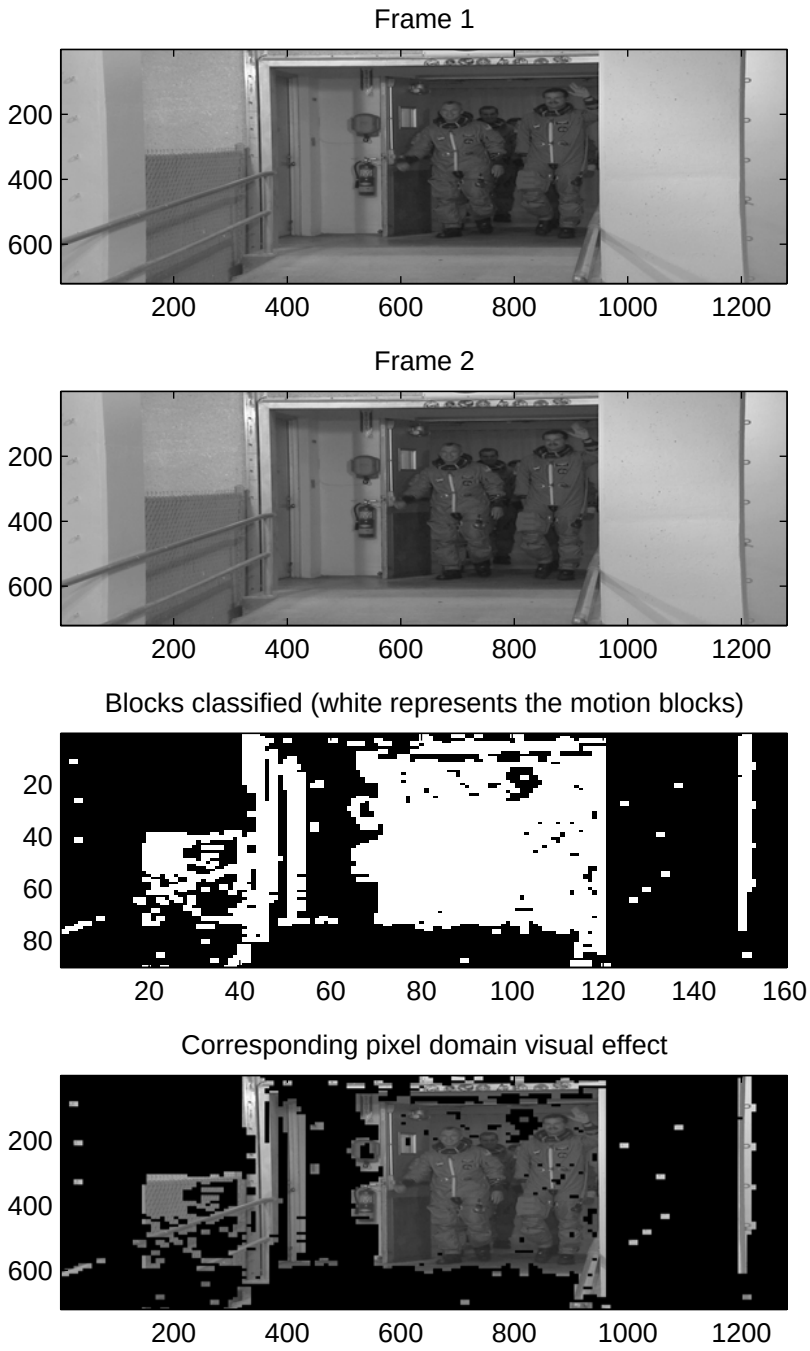


Figure 6.6: Demo 1 of the block classification (*Crew*)

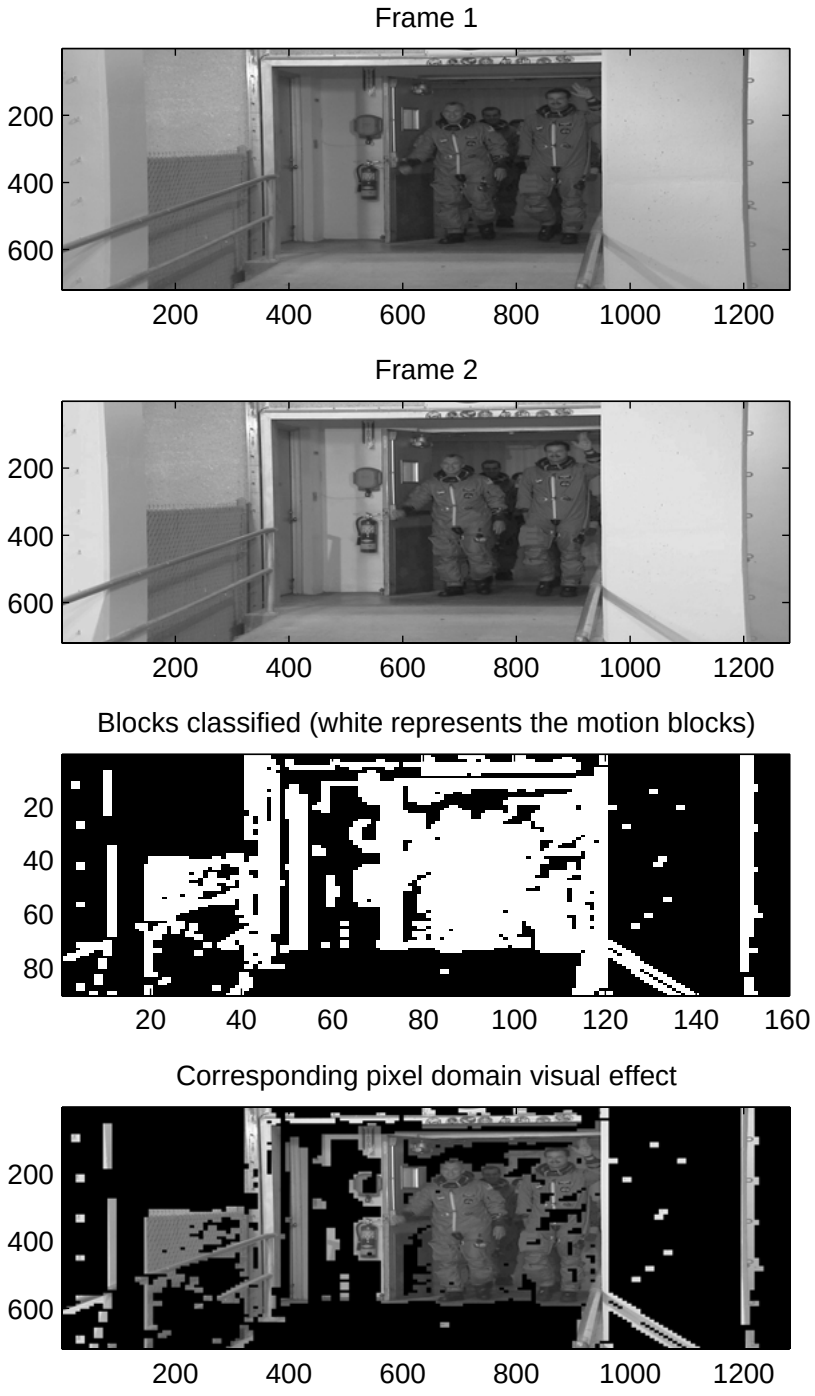


Figure 6.7: Demo 2 of the block classification (*Crew*)

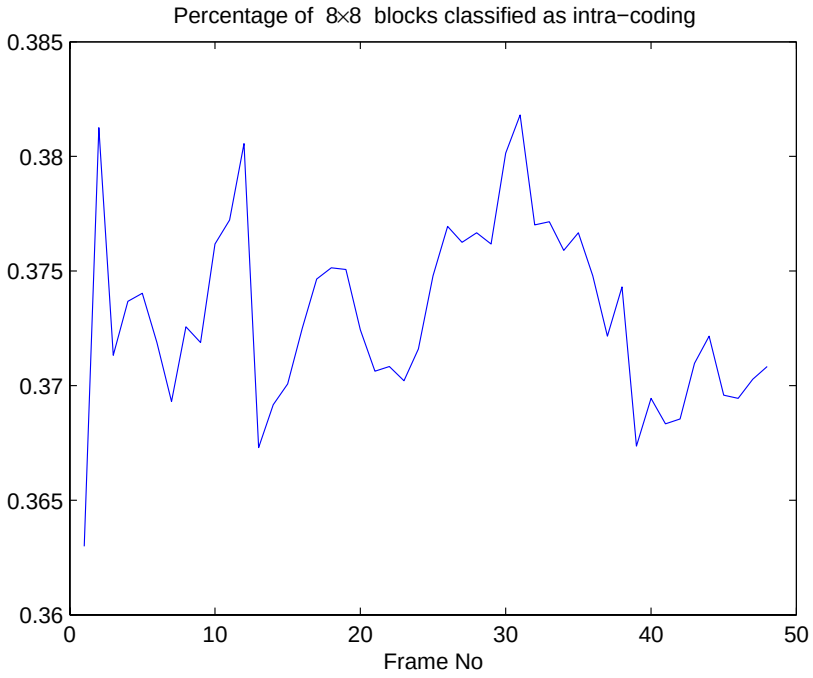


Figure 6.8: Percentage of 8×8 blocks classified as intra coding (*Crew*)

flashing effect is equivalent to scene changes, and therefore, reduces the correlation between two frames. As a consequence, the number of blocks that are intra-coded increases as compared with the previous frame.

We then test the compression performance using the proposed strategy. The motion BPE coder in chapter 5 and H.263+ obtained from (129) are used as reference coders. H.263+ is a low bit rate video coding algorithm. For H.263+, the frame sequence is set to IPPPI, i.e, the coding is one intra-coded frame followed by three predictive frames. For the BPE-DVC each key frame is followed by three frames coded by BPE-DVC.

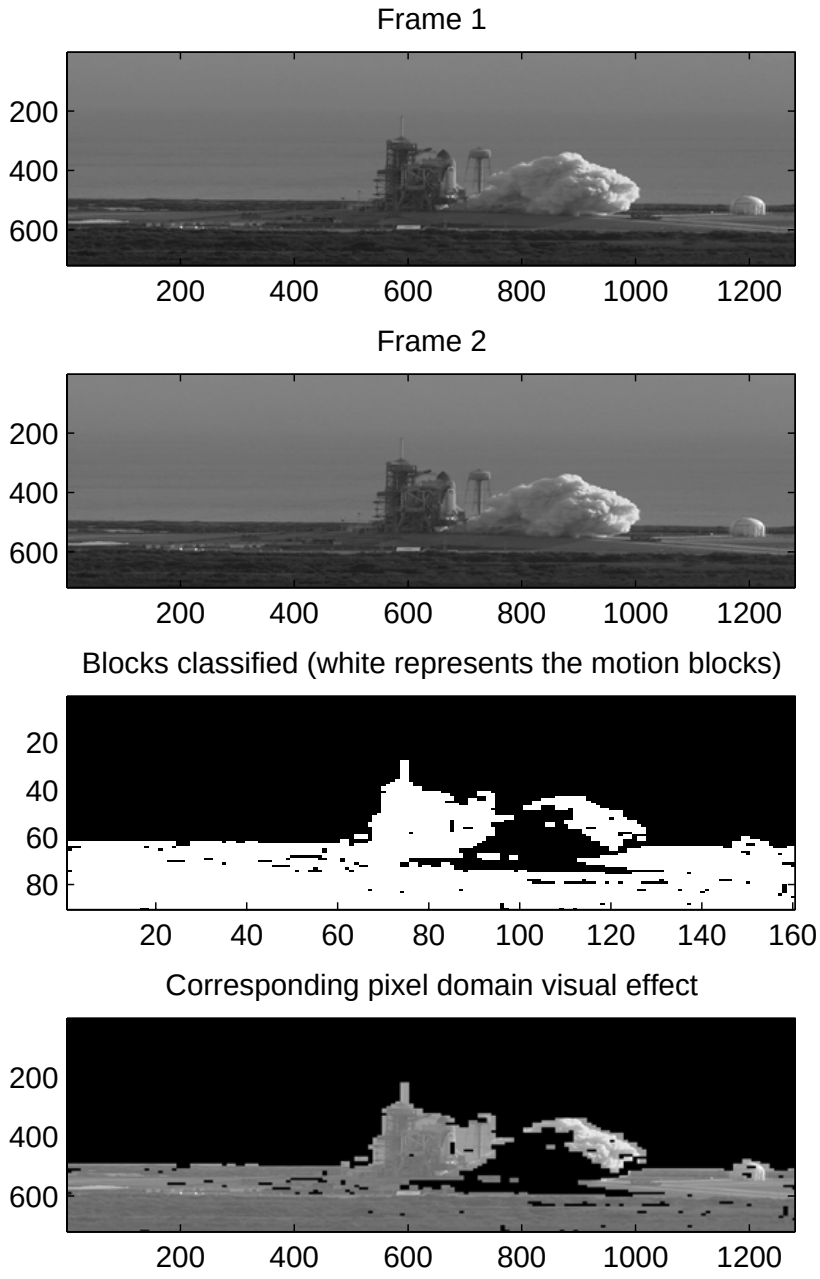


Figure 6.9: Demo 1 of the block classification (*Shuttle*)

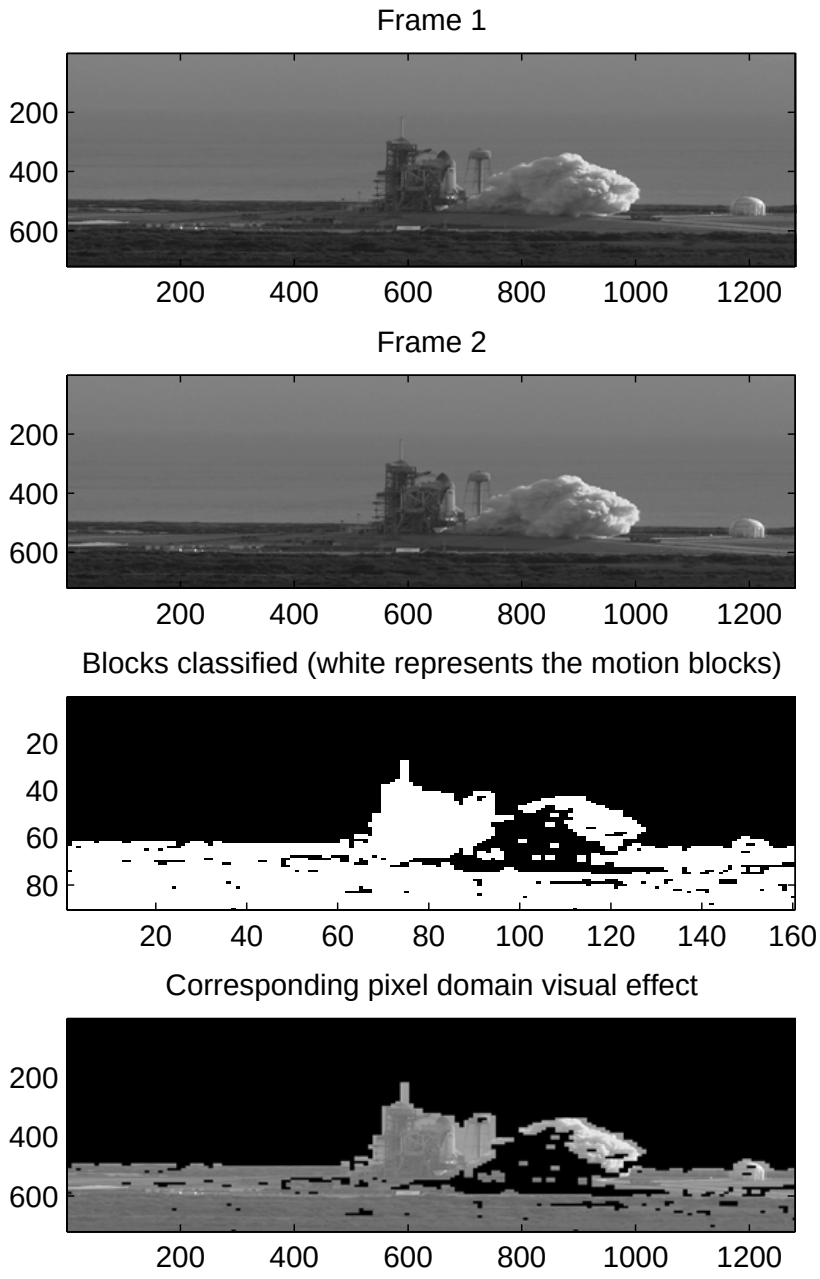


Figure 6.10: Demo 2 of the block classification (*Shuttle*)

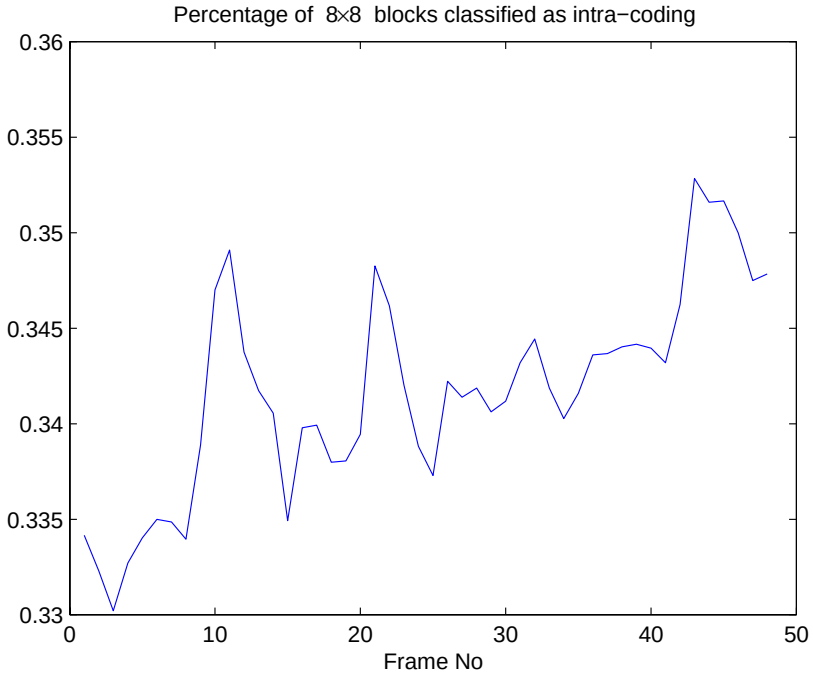


Figure 6.11: Percentage of 8×8 blocks classified as intra coding (*Shuttle*)

Figures 6.12 and 6.13 show the compression performance of intra-frame BPE coder, BPE-DVC system, and H.263+ over the 48 frames of the *Crew* and *Shuttle* sequences, respectively. We can see that the distributed video coding system can generate superior compression performance over pure intra-frame video coding, and the average compression gain over the intra-frame BPE coder is around 1.5-2.5dB. It also shows that there is still a performance gap (1.5-3dB) between the motion compensated prediction-based H.263+ and the proposed scheme. However, the encoder complexity of the proposed system is significantly lower than that of H.263+, as motion estimation and motion compensa-

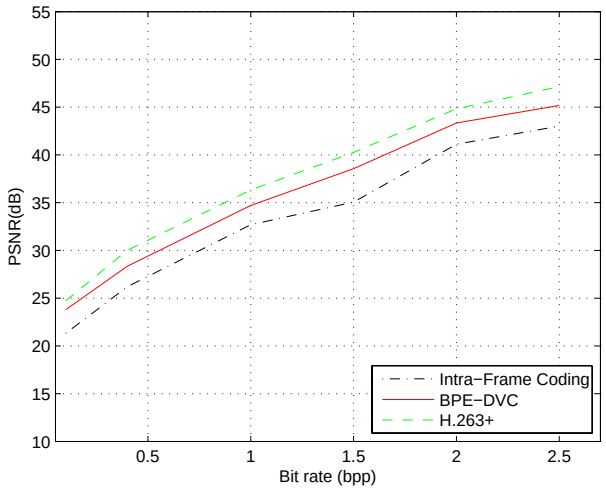


Figure 6.12: Compression performance comparison (*Crew*)

tion are avoided. The compromise of coding efficiency may be tolerable for the coding applications where encoder complexity is a serious concern.

6.4.2 Robustness Performance

Robustness of BPE-DVC is examined and compared with H.263+ under error-prone conditions. To make the comparison fair and meaningful, for BPE-DVC and H.263+, we assume that the packet headers are well protected against bit errors. This is a reasonable assumption as the packet headers usually deliver the most important information, which generally is protected by powerful error control coding. In addition, for both coders, the decoding of a packet stops when there is an invalid codeword

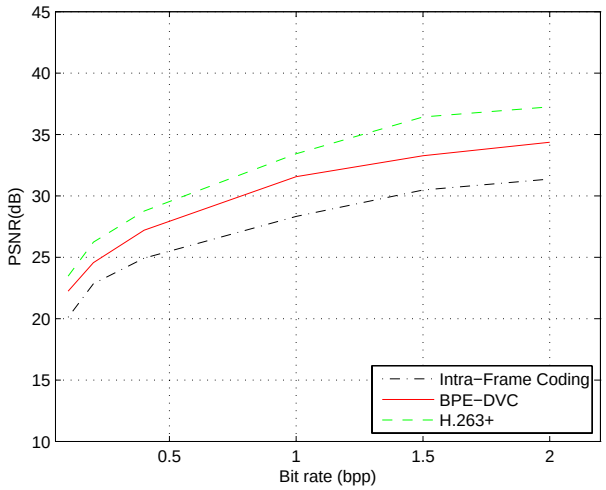


Figure 6.13: Compression performance comparison (*Shuttle*)

detected, and the packet is reconstructed based on codewords that have been decoded successfully. Assume the number of bits contained in the headers are known, the decoder skips the rest of the bits in the packet and proceed to the next packet. We also assume that the packets are always synchronized, i.e., there is no loss of bits, though bit flip errors may occur. This is to ensure that all packets are properly addressed.

The same frame structure as the one in compression performance test is used, i.e., three frames are coded based on DVC-BPE after one key frame. For BPE-DVC, the intra-coded block containing errors will not propagate to the next frame. If a block that acts as the side information for the decoding of the co-located block contains errors, the decoder performs error concealment by replicating the neighboring block, and

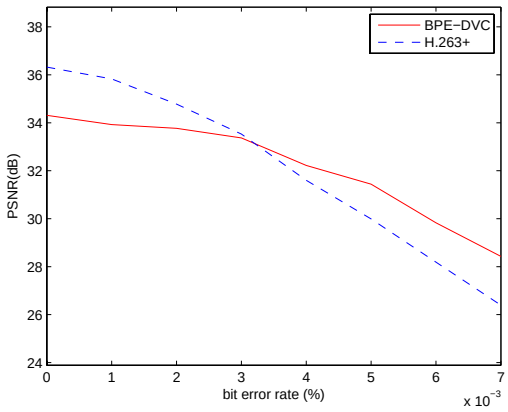


Figure 6.14: Robustness test performance(*Crew*)

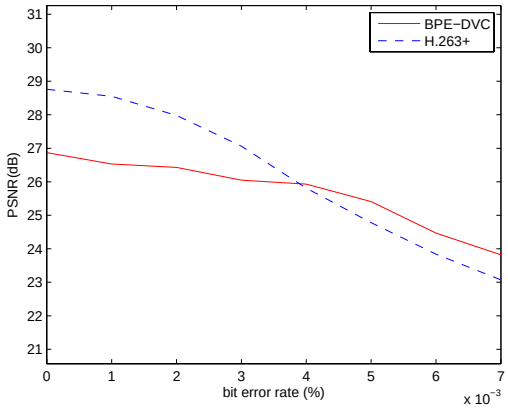


Figure 6.15: Robustness test(*Shuttle*)

the reconstructed block acts as the side information for decoding.

Figure 6.14 and 6.15 shows the simulation results of robustness performance test of BPE-DVC, and H.263+ for the 48 frames of the *Crew* and *Shuttle* sequences, respectively at 1 bit/pixel (note that the bit rate is converted from bits/second to bits/pixels, the sequence is assumed to be 30 frames/second). As we can see, BPE-DVC demonstrates more grace-

ful degradation with worsening bit error rates than H.263+, although for no error and low error rate conditions the PSNR is lower.

Robustness of BPE-DVC system is achieved from several aspects: one is that the classification of motion and still blocks improves robustness for the regions that contain the most activity and significant information. The second is attributed to the bit plane coding of the modulo value in BPE-DVC. Assume that in a segment the top two bit planes are correctly decoded, and the decoding of the third bit plane fails due to bit errors, based on the side information, the original symbols can be roughly recovered, and the error propagation and drifting is limited within the range of the side information. In addition, if bit errors occur in the refinement parts, they have no serious impact on the decoding process because they are exactly output without entropy coding. However, for the predictive coding such as H.263+, as invalid codewords interrupt the decoding process of a packet, the remaining codewords are lost. So for predictive coders like H.263+, serious error propagation occurs.

6.5 Conclusion

In this chapter, video coding is developed by exploring the inter-frame redundancy in a way that no explicit inter-frame video coding is used. As

long as statistical dependency exists, two sources can be coded independently, while recovered by exploiting the dependencies by the decoder. With block classification based on the correlation models between neighboring frames, the concept of distributed video coding is developed for low-complexity and efficient coding in the wavelet domain.

Results show that in error free conditions, BPE-DVC achieves decent compression performance. Its rate-distortion performance is superior over intra-frame based BPE coder. Though there is a coding gap to reach the rate distortion performance of motion-compensated predictive coders such as H.263+, the complexity of BPE-DVC is significantly reduced, and its robustness is significantly improved.

Chapter 7

Conclusions and Recommendations for Future Work

In this chapter, we summarize the principal contribution of the dissertation. The proposed rate allocation scheme and robust video coding schemes have potential applications in image and video applications where complexity and computational resources are limited. We also present the future research direction on these topics.

7.1 Conclusions and Contributions

The conventional rate allocation scheme used in wavelet based image coder JPEG2000 employs a post-completion rate allocation algorithm (the PCRD-Opt algorithm) to allocate bit resources and optimize the rate distortion performance (21; 22). Because the encoder has to complete the entire coding process, this method is not efficient and may not

be the best choice for applications where the exhaustive coding methods are not feasible. The model based ρ domain analysis based rate control turns out to be very accurate in rate estimate. However, the ρ domain based rate allocation has been proved to be not very accurate.

The conventional motion-compensated inter-frame video coders are generally sensitive to errors in the bitstream caused by disturbance and imperfect channel conditions. The error propagation can be catastrophic and have a very severe impact on the quality of reconstructed video sequences. To improve the robustness, and avoid computationally heavy motion compensation searches, intra-frame video coding schemes that use existing image coders as the basis for video compression have been widely used. Featuring low complexity, low computational costs, and good robustness, however, their PSNR performance is generally inferior to that of inter-frame video coding. A good balance of PSNR and robustness in video coding is desired.

To address these issues, the objective of this dissertation has been to develop rate allocation schemes for the wavelet based image coders and the intra-frame wavelet based video coders, and enhance the robustness using the concept of the distributed video coding. The wavelet based BPE coder is used for demonstration of the proposed coding schemes

while their application may be extended to other wavelet transform based image and video coders.

First we attempt to improve the rate distortion performance of the wavelet image coder by properly allocating bit resources to different coding segments. The ρ domain analysis has been proposed to incorporate the conventional PCRD-Opt algorithm. Despite its excellent rate allocation performance, the PCRD-Opt algorithm imposes high requirements on computational cost and complexity, which limit its practical applications. Combining the ρ domain rate estimation method with the PCRD-Opt algorithm, the proposed new rate allocation scheme has been applied to the segment level rate allocation. Our experiments show that the overall PSNR performance has been dramatically improved as compared with the constant bit rate allocation schemes, while the computational and complexity load has been considerably reduced as compared with the PCRD-opt algorithm.

We proposed an image classification scheme in the wavelet transform domain to classify images into *Region-of-Interest* and *Region-of-non-Interest* based on the activity level of the blocks contained in the regions. A portion of the work such as the rate allocation and significate block map are used in the robust video coding scheme proposed later.

7. CONCLUSIONS AND RECOMMENDATIONS FOR FUTURE WORK

The objective of this work is to distinguish different regions according to their activities and allocate bit resource using the proposed rate allocation method. An iterative region growing method in wavelet transform domain has been proposed. A significant block map is transmitted and the corresponding syntax elements in the header are added to indicate whether the associated blocks are in region of interest or non-interest. We show that competitive compression performance is achieved using the proposed scheme. Robustness can be enhanced if the region of interest is transmitted with higher priorities.

We then extended the image coder to video coding, and developed rate allocation algorithms for the intra-frame wavelet image coder based video coding. The rate allocation ρ-PCRD developed for image coder has been extended to frame level as well as segmentation level. A two-stage bit allocation algorithm using the previous coded frames is proposed. Our experiences show that superior compression performance has been obtained as compared with the case where each individual frame is independently coded.

After the image coder has been extended for video coding, we attempted to enhance the robustness of the coder. We examined the correlation of the neighboring frames and proposed simple and efficient robust

coding schemes based on the principal of distributed source coding. To prevent error propagation from one frame to the next on the decoding side, 8×8 blocks are classified into motion and still blocks. The motion blocks are better protected against errors using intra-frame video coding while the still blocks are coded with modulo based distributed source coding. The scheme is intended to achieve a good compromise between robustness of intra-frame video coding and compression efficiency of inter-frame video coding. Our experiments show that the scheme can yield good rate-distortion performance as compared with the intra-frame video coding based on the wavelet image coder. In addition, as compared with the H.263, robustness has been significantly increased at high bit rate errors.

7.2 Recommendation for Future Work

7.2.1 Weighted ρ Domain Based Rate Allocation

The human visual system (HVS) reveals the visual sensitivity of human being for certain types of frequency and composition signals. To incorporate it with the objective distortion measures has been an active research area.

For the ρ domain rate allocation, we do not consider the visual effect when calculating the rate distortion deduction. An image reconstructed with greater PSNR is generally considered to have better reconstruction quality. However, better PSNR may not always correlate with better visual quality because the perceived quality of an image using the human visual system (HVS) is not always consistent with PSNR measures. Distortion measures that combine objective metrics like PSNR with HVS models have been extensively studied. Since different frequency components have different impact on visual observation, rate allocation taking the HVS into consideration could lead to better visual quality, though it may not be consistent with the objective measures such as PSNR. It would be interesting to find and resolve the discrepancy such that more pleasant visual coding is achieved.

7.2.2 Distributed Source Coding for Hyperspectral Image Compression

A hyperspectral image is a special type of image that has many kinds of civilian and military applications. Unlike color images that usually have only three color spectral components, hyperspectral imaginary generally has a very large number of spectral bands, and the compression of hy-

perspectral images has been an active research topic in the past decade (130; 131; 132; 133; 134; 135; 136; 137; 138; 139; 140).

One particular interesting property in hyperspectral image is that the correlation is high between neighboring spectral bands. This scenario, as opposed to video sequences that contains motion and scene changes, is very suitable to distributed source coding. As the sequence is highly correlated, the side information could be easily determined by examining the neighboring frames. The distributed video coding performance can be expected to approach closely that of the conventional inter-band hyperspectral coding.

As a low-complexity wavelet based image coder, the BPE coder was actually developed for the coding of space data, especially for fast and efficient compression of hyperspectral images. It would be interesting to apply the BPE coder to encode the spectral bands based on the distributed source coding.

7.2.3 Distributed Source Coding for Multiview Video Coding (MVC)

Multiview video coding (MVC) is used to compress video sequences recorded simultaneously from multiple views. MVC can be used for

TV-on-demand, the next generation TV, and live video streaming on the internet and cable. Due to the huge of amount data, scalable and efficient wavelet-based video codes have been proposed (141; 142).

In MVC there are new challenges arising as compared with conventional video coding. Not only are the video sequences from one view highly correlated in the temporal domain, but the video sequences from different views may have high correlation. Therefore, besides motion estimation and motion compensation in the temporal domain for a single video sequence, a new compensation between different video sequences from different views, so called disparity compensation, is defined.

To achieve better compression for a given frame we need to search for the motion in both temporal domain and the disparity between different views. As the coding dependency exists in not only the temporal domain, but in different views, it would be a challenging problem to combine these factors and develop an optimum bit allocation mechanism for different views, subject to bit rate constraint. In other words, the problem is how to incorporate the disparity compensation into the existing rate control mechanism.

As correlation may exist between different views, if an accurate correlation model can be constructed, the distributed video coding concept

can be used for low complexity coding. Unlike video coding, there may be no communication between different views. In addition, there is view correlation which does not exist in video sequences. Therefore, correlation model construction and side information generation are more challenging than DVC for regular video sequences.

7. CONCLUSIONS AND RECOMMENDATIONS FOR FUTURE WORK

References

[1] A. Aaron and B. Girod. **Compression with side information using Turbo codes**. *Proc. of IEEE Data Compression Conference*, pages 252 – 261, April 2002. ix, 35, 38, 118

[2] R.salami C.Laflammw J.-P. Adoul and D.Massalopux. **A toll quality 8 kb/s speech codec for the personal communications system (pcs)**. *IEEE Trans. Vehic. Techn.*, **43**(3):808–816, 1994. 2

[3] ITU-T Recommendation. **G.722.1, Coding at 24 and 32 kbit/s for hands-free operation in systems with low frame loss**, 1999. 2

[4] C. Evci. **Speech Codec Aspects for Third Generation Mobile Systems**. *Proc. IEEE Veh.Tech. Conf.*, pages 172–175, 1992. 2

[5] D. W. Petr. **32 kb/s ADPCM-DLQ coding for network applications**. *Proc. IEEE Global Telecomm. Conf.*, pages A8.3–1 – A8.3–5, 1982. 2

[6] W. R. Daumer, X. Maitre, P. Mermelstein, , and I. Tokizawa. **Overview of the 32 kb/s ADPCM algorithm**. *Proc. IEEE Global Telecomm. Conf.*, pages 791–795, 1984. 2

[7] P. Mermelstein. **G.722, a new ccitt coding standard for digital transmission of wideband audio signals**. *IEEE Comm. Mag.*, **26**(1):8–15, 1988. 2

[8] N. S. Jayant and Peter Noll. *Digital Coding of waveforms, principles and applications to speech and video*. Prentice-Hall, 1984. 2

[9] JPEG committee. **ISO 10918-1 Information technology- Digital compression and coding of continuous-tone still images**. 1994. 2

[10] ITU-T. **ISO/IEC 11544:1993 Information technology - Coded representation of picture and audio information – Progressive bi-level image compression**. 1994. 2

[11] ISO/IEC. **ISO/IEC 15444-1:2004 Information technology – JPEG 2000 image coding system: Core coding system**, 2000. 2, 13

[12] M. J. Weinberger, G. Seroussi, and G. Sapiro. **The LOCO-I lossless image compression algorithm: Principles and standardization into JPEG-LS**. *IEEE Trans. Image Process*, **9**:1309–1324, August 2000. 2

[13] ITU-T. **ITU-T Recommendation H.261 version 2, Video codec for audiovisual services at p× 64k**. 1990. 2, 30

[14] ITU-T Recommendation. **H.263, Video coding for low bit rate communication**, 1998. 2, 30

[15] ITU-T. **ISO/IEC 11172-2:1991, Coding of moving pictures and associated audio for digital storage media at up to about 1.5 Mbps, Part 2: Visual**. 1991. 2

[16] ITU-T. **ISO/IEC 13818-2:1994, Information technology-generic coding of moving pictures and associated audio, Part 2**. 1994. 2

[17] ITU-T. **ISO/IEC 14496-2:1999, Information technologycoding of audio/visual objects, Part 2, Visual**. 1999. 2

[18] ITU-T. **H.264: Advanced video coding for generic audiovisual services**. 2003. 2

[19] http://www.itu.int/ITU T/studygroups/com16/sg16-q6.html. **ITU-T Study Group**. 2

[20] Witness the development of H.265. http://www.h265.net/. 2

[21] David S. Taubman and Michael W. Marcellin. *JPEG2000: Image Compression Fundamentals, Standards, and Practice*. Springer, 2002. 3, 22, 23, 55, 57, 58, 145

[22] M. Smith and J. Villasenor. **Intra-Frame JPEG2000 vs Inter-frame Compression Comparison: The benefits and Trade-offs for Very high Quality , HIgh-resolution Sequences**. Pasadena, CA, October 2004. SMPTE Technical Conference. 3, 102, 104, 145

[23] Image Data Compression. **Recommendation for Space Data System Standards, Blue Book**. *CCSDS 122.0-B-1*, **1**, November 2005. 4, 13, 39, 41

[24] Zhihai He. *ρ domain rate-distortion analysis and rate control for visual coding and communication*. PhD thesis, University of California at Santa Barbara, 2001. 4, 58, 59, 64

[25] Zhihai He and S.K. Mitra. **A linear source model and a unified rate control algorithm for DCT video coding**. *Circuits and Systems for Video Technology, IEEE Transactions on*, **12**(11):970–982, November 2002. 4, 58

[26] Zhihai He and S. K. Mitra. **Optimum bit allocation and accurate rate control for video coding via /spl rho/-domain source modeling**. *Circuits and Systems for Video Technology, IEEE Transactions on*, **12**(10):840–849, October 2002. 4, 28, 58

[27] He Z and Wu D. O. **Linear Rate Control and Optimum Statistical Multiplexing for H.264 Video Broadcast**. *Multimedia, IEEE Transactions on*, **10**(7):1237 – 1249, November 2008. 4

[28] Slepian and J.K. Wolf. **Noiseless coding of correlated information sources**. *IEEE Trans. Inform. Theory*, **19**(4):471–480, November 1973. 6, 34, 116

[29] A. Wyner and J. Ziv. **The rate-distortion function for source coding with side information at the decoder**. *IEEE Trans. Inform. Theory*, **22**(1):1–10, January 1976. 6, 34, 116

[30] H. S. Malavar. *Signal Processing with Lapped Transforms*. Artech House, Norwood, MA, 1992. 13

[31] Donald B. Percival and Andrew T. Walden. *Wavelet Methods for Time Series Analysis*. Cambridge University Press, 2000. 13

[32] Paul S. Addison. *The Illustrated Wavelet Transform Handbook*. Institute of Physics, 2002. 13

[33] Shapiro J. M. **embedded image coding using zerotrees of wavelet coefficients**. *IEEE Transactions on Signal Processing*, **41**(1):3445–3462, June 1993. 13

[34] Amir Said and William A. Pearlman. **A New Fast and Efficient Image Codec Based on Set Partitioning in Hierarchical Trees**. *IEEE Transactions on Circuits and Systems for Video Technology*, **6**(1):243–250, June 1996. 13

[35] D. Taubman. **High performance scalable image compression with EBCOT**. *Image Processing, IEEE Transactions on*, **9**(7):1158 – 1170, July 2000. 13

[36] Amir Said and William A. Pearlman. **Image Compression with Set Partitioning in Hierarchical Trees**. http://www.cipr.rpi.edu/research/SPIHT/spiht0.html, June 2006. 14, 22

[37] M. Vetterli and J. Kovacevic. *Wavelets and Subband Coding*. Prentice Hall, 1995. 15

[38] K. Sayood. *Introduction to Data Compression*. Kaufmann, San Mateo, CA, 2000. 17, 22

[39] Beong-Jo Kim and W.A. Pearlman. **An embedded wavelet video coder using three-dimensional set partitioning in hierarchical trees(SPIHT)**. **41**, pages 251– 260, Utah, March 1997. Data Compression Conference. 19, 20

[40] S.A. Martucci, I. Sodagar, T. Chiang, and Ya-Qin Zhang. **A zerotree wavelet video coder**. *IEEE Transactions on Circuits and Systems for Video Technology*, **7**(1):109–118, February 1997. 19, 20

[41] Lazar D. and Averbuch A. **Wavelet-based video coder via bit allocation**. *Circuits and Systems for Video Technology IEEE Transactions on*, **11**(7):815 – 832, July 2001. 19, 28

[42] K.K. Lin and R.M. Gray. **Wavelet video coding with dependent optimization**. *Circuits and Systems for Video Technology, IEEE Transactions on,*, **14**(4):542–553, April 2004. 19, 33

[43] D.S. Turaga, van der Schaar M., and B. Pesquet-Popescu. **Complexity scalable motion compensated wavelet video encoding**. *Circuits and Systems for Video Technology, IEEE Transactions on*, **15**:982 – 993, August 2005. 19

[44] Ruiqin Xiong, Jizheng Xu, Feng Wu, Shipeng Li, and Ya-Qin Zhang. **Subband Coupling Aware Rate Allocation for Spatial Scalability in 3-D Wavelet Video Coding**. *Circuits and Systems for Video Technology, IEEE Transactions on*, **17**(10):1311–1324, October 2007. 19, 28

[45] Liu Y., Wu F., and Ngan K. N. **3-D Object-Based Scalable Wavelet Video Coding With Boundary Effect Suppression**. *Circuits and Systems for Video Technology, IEEE Transactions on*, **17**(5):639 – 644, May 2007. 19

[46] K. Cinkler. **Very low bit-rate wavelet video coding**. *Selected Areas in Communications, IEEE Journal on*, **16**(1):4 – 11, January 1998. 19

[47] Fuzheng Yang, Shuai Wan, Yilin Chang, and Hong Ren Wu. **A novel objective no-reference metric for digital video quality assessment**. *Signal Processing Letters,IEEE*, **12**(10):685 – 688, October 2005. 21

[48] EePing Ong, Weisi Lin, Zhongkang Lu, Susu Yao, and M. Etoh. **Visual distortion assessment with emphasis on spatially transitional regions**. *Circuits and Systems for Video Technology, IEEE Transactions on*, **14**(4):559 – 566, April 2004. 21

[49] J. Oh, S.I. Woolley, T.N. Arvanitis, and J.N. Townend. **A multistage perceptual quality assessment for compressed digital angiogram images**. *Medical Imaging, IEEE Transactions on*, **20**(12):1352 – 1361, December 2001. 21

[50] K.A Panetta, E.J. Wharton, and S.S. Agaian. **Human Visual System-Based Image Enhancement and Logarithmic Contrast Measure**. *Systems, Man, and Cybernetics, Part B, IEEE Transactions on*, **38**(1):174 – 188, February 2008. 22

[51] Image Data Compression. **Recommendation for Space Data System Standards, Green Book**. *CCSDS 122.0-G-1*, **1**, June 2007. 23, 39, 41, 73, 76

[52] T. Berger. *Rate Distortion Theory.* Prentice-Hall, 1971. 23

[53] L. Overmeire, L. Nachtergaele, F. Verdicchio, J. Barbarien, and P. Schelkens. **Constant quality video coding using video content analysis**. *Signal Processing: Image Communications*, **20**:343369, 2005. 28, 30

[54] A. Ortega, K. Ramchandran, and M. Vetterli. **Optimal trellisbased buffered compression and fast approximations**. *IEEE Trans. Image Process*, (3):2640, 1994. 28

[55] K. Ramchandran, A. Ortega, and M. Vetterli. **Bit allocation for dependent quantization with applications to multiresolution and MPEG video coders**. *IEEE Trans. Image Process.*, (3):533545, 1994. 28

[56] G.M. Schuster, G. Melnikov, and A.K. Katsaggelos. **A review of the minimum maximum criterion for optimal bit allocation among dependent quantizers**. *IEEE Trans. Multimedia*, page 317, 1999. 28

[57] H. Everett. **Generlized Lagrange multiplier method for solving problems of optimum allocation of resource**. *Oper. Res.*, **11**:399417, 1963. 28

[58] A. Ortega and K. Ramchandran. **Rate-distrotion methods for image and video compression**. *IEEE Signal Process. Magazine*, 1998. 30

[59] G.M. Schuster and A.K. Katsaggelos. *Rate Distortion based Video Compression.* Kluwer Academic Publishers, 1997. 30

[60] E.C. Reed and J.S. Lim. **Optimal multidimensional bit-rate control for video communications**. *IEEE Trans. Image Process*, page 873885, November 2002. 30

[61] CCITT. **CCITT SG XV WP/1/Q4, Description of reference model 8 (RM8)**. 1989. 30

[62] ITU-T JVT committee. **Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG, Draft ITU-T recommendation and final draft international standard of joint**

video specification (ITU-T Rec. H.264 ISO/IEC 14496-10 AVC), JVT-G050. 2003. 30

[63] Wiegand T., Sullivan G.J., Bjontegaard G., and Luthra A. **Overview of the H.264/AVC video coding standard**. *Circuits and Systems for Video Technology, IEEE Transactions on*, **13**:560–576, July 2003. 30

[64] G. Sullivan, T. Wiegand, and K.-P. Lim. **Joint model reference encoding methods and decoding concealment methods, JVT-I049**. September 2003. 30

[65] Po-Yuen Cheng, Jin Li, and Jay Kuo C.-C. **Rate control for an embedded wavelet video coder**. *IEEE Transactions on Circuits and Systems for Video Technology*, **7**(4):696–702, August 1997. 33

[66] Mingshi Wang and M. van der Schaar. **Operational rate-distortion modeling for wavelet video coders Signal Processing**. *Signal Processing, IEEE Transactions on*, **54**(9):3505 – 3517, September 2006. 33

[67] Chao He, Jianyu Dong, Zheng Y.F., and Zhigang Gao. **Optimal 3-D coefficient tree structure for 3-D wavelet video coding**. *Circuits and Systems for Video Technology, IEEE Transactions on*, **13**(10):961 – 972, 2003. 33

[68] Ruiqin Xiong, Jizheng Xu, Feng Wu, and Shipeng Li;. **Barbell-Lifting Based 3-D Wavelet Coding Scheme**. *Circuits and Systems for Video Technology, IEEE Transactions on*, **17**(9):1256 – 1269, 2007. 33

[69] Tillier C., Pesquet-Popescu, and B. van der Schaar M. **3-band motion-compensated temporal structures for scalable video coding**. *Image Processing, IEEE Transactions on*, **15**(9):2545 – 2557, September 2006. 33

[70] Balter R., Gioia P., and Morin L. **Scalable and Efficient Video Coding Using 3-D Modeling**. *Multimedia, IEEE Transactions on*, **8**(6):1147 – 1155, December 2996. 33

[71] S.S. Pradhan and K. Ramchandran. **Distributed source coding using syndromes (DISCUS): design and construction**. *IEEE Trans. Inform. Theory*, **49**(3):626–643, March 2003. 35

[72] J. Garcia-Frias and Y. Zhao. **Compression of correlated binary sources using Turbo codes**. *IEEE Commun. Letters*, **21**(5):417–419, October 2001. 35

[73] J. Bajcsy and P. Mitran. **Coding for the Slepian-Wolf problem with Turbo codes**. *Proc. IEEE GLOBECOM*, **21**(5):417–419, November 2001. 35, 38

[74] J. Li, Z. Tu, and R. S. Blum. **Slepian-Wolf coding for nonuniform sources using Turbo codes**. *Proc. IEEE Data Compression Conf.*, pages 312–321, March 2003. 35

[75] A. D. Liveris, Z. Xiong, and C. N. Georghiades. **Distributed compression of binary sources using conventional parallel and serial concatenated convolutional codes**. *Proc. of IEEE Data Compression Conference (DCC)*, pages 193–202, March 2003. 35

[76] Y. Matsunaga and H. Yamamoto. **A coding theorem for lossy data compression by LDPC codes**. *Information Theory, IEEE Transactions*, **49**(9):2225 – 2229, September 2003. 35

[77] M. Sartipi and F. Fekri. **Source and channel coding in wireless sensor networks using LDPC codes**. *EEE SECON 2004. 2004 First Annual IEEE Communications Society Conference on*, **49**:309 – 316, October 2004. 35, 38

[78] Yang Yang, S. Cheng, Zixiang Xiong, and Wei Zhao. **Wyner-Ziv coding based on TCQ and LDPC codes**. *Conference Record of the Thirty-Seventh Asilomar Conference on*, **1**:825 – 829, November 2003. 35

[79] A.D. Liveris and Zixiang Xiong. **Compression of binary sources with side information at the decoder using LDPC codes**. *Georghiades,C.N.; Communications Letters, IEEE*, **6**(10):440 –442, October 2002. 35

[80] B. Girod, A.M. Aaron, S. Rane, and D. Rebollo-Monedero. **Distributed video coding**. *Proc. Asilomar Conf. Signals, Syst., Comput.*, **93**(1):71–83, January 2005. 35

[81] Shu Lin and JR. Daniel J. Costello. *Error control coding.* Prentice Hall, 2004. 36

[82] C. Berrou, A. Glavieux, and P. Thitimajshima. **Near Shannon Limit Error-Correcting Coding: Turbo Codes**. *Proc. IEEE Int. Conf. on Commun.*, pages 1064–1070, May 1993. 36

[83] Todd K. Moon. *Error Correction Coding, Mathematical Methods and Algorithms.* Wiley, 2005. 36

[84] Bernard Sklar. *Digital communication, fundamentals and applications.* Prentice-Hall, 2001. 36

[85] J. Roberto and B. de Marca. **On Noisy Channel Quantizer Design for Unequal Error Protection**. *Speech and Audio Coding for Wireless and Network Applications, Kulwer Academic Publishers*, 1993. 37

[86] Joohee Kim, Mersereau R.M., and Altunbasak Y. **Error-resilient image and video transmission over the Internet using unequal error protection**. *Image Processing, IEEE Transactions on*, **12**(2):121 – 131, February 2003. 37

[87] Joohee Kim, Mersereau R.M., and Y. Altunbasak. **Distributed video streaming using multiple description coding and unequal error protection**. *Image Processing, IEEE Transactions on*, **14**(7):849 – 861, July 2005. 37, 38

[88] Miska M. Hannuksela, Ye kui Wang, and Moncef Gabbouj. **Sub-picture video coding for unequal error protection**. XI European Signal Processing Conference, 2002. 37

[89] Y. Takishima, M. Wada, and H. Murakami. **Reversible variable length codes**. *IEEE Trans. Commun.*, **43**, 1995. 37

[90] Chie-Wu Tsai and Ja-Ling Wu. **On constructing the Huffman-coded-based reversible variable length codes**. *IEEE Trans. Commun.*, **49**(9), 2001. 37

[91] C.-W. Tsai and J.-L. Wu. **A modified symmetrical reversible variable length code and its theoretical bounds**. *IEEE Trans. Inform. Theory.*, **47**:2543–2548, September 2001. 37

[92] J. Wen and J. D. Villasenor. **A class of reversible variable length codes for robust image and video coding**. *Proc. IEEE Int. Conf. Image Processing*, **2**:65–68, 1997. 37

[93] J. Wen and J. D. Villasenor. **Reversible variable length codes for efficient and robust image and video coding**. *Data Compression Conf.*, pages 471–480, 1998. 37

[94] M. Titchener. **The synchronization of variable length codes**. *IEEE Trans. Inform. Theory*, **43**(2):683–691, 1997. 37

[95] U. Gunther. *Robust source coding with generalized T-codes*. PhD thesis, University of Auckland, Auckland, 1998. 37

[96] M. Titchener. **Generalised T-codes: extended construction algorithm for self-synchronising codes**. *IEE Proc.-Commun.*, **143**(122-128), 1996. 37

[97] M. Titchener. **Digital encoding by means of new T-codes to provide improved data synchronization and message integrity**. *IEE Proc. Comput Digital Tech*, **131**(4):151–153, 1984. 37

[98] C. E. Shannon. **A mathematical theory of communication**. *Bell Syst.Tech. J.*, **27**:379–423–623–656, 1948. 37

[99] T. C. Ancheta Jr. *Joint source channel coding*. PhD thesis, Univ. Notre Dame, August 1977. 37

[100] J. L. Massey. **Joint source and channel coding**. *Communication Systems and Random Process Theory*, pages 279–293, 1978. 37

[101] S. Vembu, S. Verdu, and Y. Steinberg. **The source channel theorem revisited**. *IEEE Trans. Inform. Theory*, **41**:44–54, January 1995. 37

[102] A. J.Kurtenbach and P. A.Wintz. **Quantizing for noisy channels,**. *IEEE Trans. Commun.*, **COM-17**:291–302, April 1969. 37

[103] K. Sayood, H.H. Otu, and N. Demir. **Joint source/channel coding for variable length codes**. *Communications, IEEE Transactions on*, **48**(5):787 – 794, May 2000. 37

[104] B. Belzer, J. D. Villasenor, and B. Girod. **Joint source channel coding of images with trellis coded quantization and convolutional codes**. *in Proc. 1995 Int. Conf. Image Processing*, pages 85–88, October 1995. 37

[105] K. Sayood and J. C. Borkenhagen. **Use of residual redundancy in the design of joint source/channel coders**. *IEEE Trans. Commun.*, **39**:838–846, March 1991. 37

[106] K. Sayood, F. Liu, and J. D. Gibson. **Constrained joint source/channel coder design**. *IEEE J. Select. Areas Commun.*, **12**:1584–1593, December 1994. 37

[107] M.Park and D. J. Miller. **Joint source-channel decoding for variable length encoded data by exact and approximate MAP sequence estimation**. *in Proc. Int. Conf. Acoustics, Speech, Signal Processing*, pages 2451–2454, March 1999. 37

[108] S.S. Hemami. **Robust video coding - an overview, Acoustics, Speech, and Signal Processing**. **5**, pages 18–23. Proceedings. (ICASSP '05). IEEE International Conference on, March 2005. 37

[109] B. Girod, A.M. Aaron, and S. Rane. **Distributed video coding**. *Proceedings of the IEEE*, **93**(1):71 – 83, January 2005. 38

[110] Zixiang Xiong, A.D. Liveris, and S. Cheng. **Distributed source coding for sensor networks**. *Signal Processing Magazine, IEEE*, **21**(5):80–94, September 2004. 38

[111] V. K. Goyal. **Multiple Description Coding: Compression Meets the Network,**. *IEEE Signal Processing Magazine*, **18**(5):74–94, September 2001. 38

[112] R. Puri and K. Ramchandran. **Multiple description source coding through forward error correction codes**. *IEEE Proceedings Asilomar Conference on Signals, Systems, and Computers*, October 1999. 38

[113] CCSDS. **Lossless data compression: CCSDS report concerning space data systems standards**. Washington, DC, 1997. 39

[114] J. Villasenor, B. Belzer, and J. Liao. **Wavelet Filter Evaluation for Image Compression**. *IEEE Transactions on Image Processing*, **2**:1053–1060, August 1995. 39

[115] M. D. Adams and F. Kossentini. **Reversible Integer-to-Integer Wavelet Transforms for Image Compression: Performance Evaluation and Analysis**. *IEEE Transactions on Image Processing*, **9**(6):1010 – 1024, June 2000. 39

[116] S. W. Golomb. **Run-length encodings**. *IEEE Trans Info Theory*, **12**, 1966. 47

[117] R. F. Rice and R. Plaunt. **Adaptive Variable-Length Coding for Efficient Compression of Spacecraft Television Data**. *IEEE Transactions on Communications*, **16(9)**:889–897, Dec 1971. 47

[118] E.Y. Lam and J.W. Goodman. **A mathematical analysis of the DCT coefficient distributions for images**. *Image Processing, IEEE Transactions on*, **9**(10):1661–1666, October 2000. 59

[119] R.W. Buccigrossi and E.P. Simoncelli. **Image compression via joint statistical characterization in the wavelet domain**. *Image Processing, IEEE Transactions on*, **8**:1688–1701, December 1999. 59

[120] M. D. Adams and F. Kossentini. **Jasper: A software-based JPEG-2000 codec implementation**. *in Proc. of IEEE ICIP Conf., Vancouver, Canada*, October 2000. 73

[121] A. Aaron, R. Zhang, and B. Girod. **Wyner-Ziv coding of motion video**. *Asilomar Conf. Signals and Systems, Pacific Grove,CA*, 2002. 117

[122] A. Aaron, R. Zhang, and B. Girod. **Wyner-Ziv coding for video: applications to compression and error resilience**. *IEEE Data Compression Conf.*, pages 93–102, March 2003. 117

[123] A. Aaron, E. Setton, and B. Girod. **Toward practical Wyner-Ziv coding of video**. *IEEE Int. Conf. Image Processing, Barcelona, Spain*, 2003. 117

[124] R. Puri, A. Majumda, and K. Ramchandran. **PRISM: A Video Coding Paradigm With Motion Estimation at the Decoder**. *Image Processing, IEEE Transactions on*, **16**(10):2436 – 2448, 10 2007. 119

[125] R. Puri, A. Majumdar, and K. Ramchandran. **Distributed video coding in wireless sensor networks**. *Signal Processing Magazine, IEEE*, **23**(4):94 – 106, Jul. 2006. 119

[126] Xun Guo, Yan Lu, Feng Wu, and Wen Gao. **Distributed video coding using wavelet**. *Circuits and Systems, 2006. ISCAS 2006. Proceedings. 2006 IEEE International Symposium on*, page 4, May 2006. 120

[127] R. Bernardini, R. Rinaldo, P. Zontone, D. Alfonso, and A. Vitali. **Wavelet Domain Distributed Coding for Video**. *Image Processing, 2006 IEEE International Conference on*, pages 245–248, October 2006. 120

[128] A. Sehgal and N. Ahuja. **Robust predictive coding and the Wyner-Ziv problem**. *Proc. IEEE Data Compression Conf.*, pages 103–112, 2003. 121

[129] Trac D. Tran. **H.263+ TMN**. Web:thanglong.ece.jhu.edu/Course/443/Lectures/H.263+/. 138

[130] F. Rizzo, B. Carpentieri, G. Motta, and J. A. Storer. **Low-complexity lossless compression of hyperspectral imagery via linear prediction**. *IEEE Signal Processing Letters*, **12**:138 – 141, February 2005. 151

[131] J. Mielikainen and P. Toivanen. **Clustered DPCM for the lossless compression of hyperspectral images**. *IEEE Trans. Geosci. Remote Sens.*, **41**:2943–2946, February 2003. 151

[132] X. Wu and N. D. Memon. **Context-based lossless interband compression-extending CALIC**. *IEEE Trans. Image Process*, **9**:994–1001, June 2000. 151

[133] B. Aiazzi, P. Alba, L. Alparone, and S. Baronti. **Lossless compression of multi/hyper-spectral imagery based on a 3-D fuzzy prediction**. *IEEE Trans. Geosci. Remote Sens.*, **37**:2287–2294, September 1999. 151

[134] R. E. Roger and M. C. Cavenor. **Lossless compression of AVIRIS images**. *IEEE Trans. Image Process*, **5**:713–719, May 1996. 151

[135] N. D. Memon, K. Sayood, and S. S Magliveras. **Lossless compression of multispectral image data**. *IEEE Trans. Geosci. Remote Sens.*, **32**:282–289, March 1994. 151

[136] M. J. Ryan and J. F. Arnold. **The lossless compression of AVIRIS images by vector quantization**. *IEEE Trans. Geosci. Remote Sens.*, **35**:546–550, May 1997. 151

[137] H. Wang and S. D Babacan and K. Sayood. **Lossless Hyperspectral Image Compression Using Context-based Conditional Averages**. In *Proc. IEEE the Data Compression Conference*, pages 479–488, March 2005. 151

[138] F. Rizzo, B. Carpentieri nad G. Motta, and J. A. Storer. **High performance compression of hyperspectral imagery with reduced search complexity in the compressed domain**. In *Proc. IEEE the Data Compression Conference*, pages 479–488, March 2004. 151

[139] G. Motta, F. Rizzo, and J. A. Storer. **Compression of hyperspectral imagery**. In *Proc. IEEE the Data Compression Conference*, pages 333–342, March 2003. 151

[140] S. Tate. **Band ordering in lossless compression of multispectral images**. In *Proc. IEEE the Data Compression Conference*, pages 311–320, March 1994. 151

[141] Hoon Kim, Garcia J., and Ortega Jae. **Dependent bit allocation in multiview video coding**. **2**, pages II – 293–96. Image Processing, 2005. ICIP 2005. IEEE International Conference on, September 2005. 152

[142] W. Yang, Y. Lu, F. Wu, J. Caiand K. N. Ngan, and S. Li. **4-D Wavelet-Based Multiview Video Coding,**. *Circuits and Systems for Video Technology, IEEE Transactions on,*, **16**(11):1385 – 1396, November 2006. 152