CONTRIBUTIONS TO IMAGE AND VIDEO CODING FOR RELIABLE AND SECURE COMMUNICATIONS

by

Phoom Sagetong

A Dissertation Presented to the
FACULTY OF THE GRADUATE SCHOOL
UNIVERSITY OF SOUTHERN CALIFORNIA
In Partial Fulfillment of the
Requirements for the Degree
DOCTOR OF PHILOSOPHY
(ELECTRICAL ENGINEERING)

August 2004

Dedication

This dissertation is dedicated to my beloved parents, Phol and Siriporn Sagetong, my sister, Vichaya Sagetong, for their encouragement, motivation, guidance, and love.

Acknowledgements

I am extremely grateful to Professor Antonio Ortega, my academic advisor and chairman of my thesis committee for his excellent guidance, patient listening for my entire problem, constant encouragement and support during the course of this research. I am also very thankful to Dr. Wensheng Zhou for her valuable suggestions for my watermarking work at HRL Laboratories as well as serving in my qualification exam committee. I would like to thank Dr. Keith M. Chugg, Dr. C.C. Jay Kuo and Dr. Roger Zimmermann for many precious suggestions and serving in my Ph.D. qualification exam committee.

I would like to take this opportunity to thank Dr. Solomon W. Golomb, Dr. Robert M. Gagliardi, Dr. P. Vijay Kumar, Dr. Joao Hespanya, Dr. Roberto Manduchi and Dr. Zhen Zhang for enjoyable grader experience and generous discussion. I thank Dr. Alexander Sawchuk for a great research discussion in the first year of my Ph.D. program. I thank the current members of Jet Propulsion Laboratory (JPL-NASA), Sam Dolinar and Matt Klimesh, and ex-member, Dr. Roberto Manduchi for wonderful discussion. I would like to thank Dr. Ryo Bo

and Dr. Wensheng Zhou for giving me a great opportunity to work at HRL Laboratories in 2001. It was my honor to have an opportunity to work with Dr. Robert Scholtz for the great experience in Ultra Wideband (UWB) research group in 1999 and the UWB members including: Dr. Jean-Marc R. Cramer, Dr. Carlos G. Corrada-Bravo and Dr. Joon-Yong Lee. I would like to thank CSI, SIPI, IMSC, DEN, OIS and EE staffs for providing the pleasant environment during my Ph.D. life at USC since 1996 especially Tim Boston, Mayumi A. Thrasher, Gerrielyn Ramos, Diane Demetras, Lisette Garcia-Miller, Gloria Halfacre, Milly Montenegro, Susan Moore, Regina Morton, Linda Varilla, Allanexit G. Weber, Linda Wright, Ray Fujioka, Isako and many others.

I would like to thank all of the former and current members in my research group for the valuable friendship and life/research discussion including the former members: Paul Fernandez, Dr. Wenqing Jiang, Dr. Krisda Lengwehasatit, Dr. Raghavendra Singh, Patrick Kehoe, Dr. Zhourong Miao, Dr. David W. Pan, Hironori Komi, Julian Cabrera, David Comas, Jose Ramon Gimeno, Marco Fumagalli, Dr. Baltasar Beferull-Lozana, Dr. Sang-Yong Lee, Hyungsuk Kim, Young-Gap Kwon, Dr. Naveen Srinivasamurthy and Dr. Kemal Demirciler and the current members: Dr. Hua Xie, Dr. Hyukjune Chung, Changsung Kim, Nazeeh Aranki, Alexandre Ciancio, Lavanya Vasudevan, Chang-sung Kim, Huisheng Wang, Bae-Sy Brian Lan, David Romacho, Hsin-yi Ivy Tseng. They all have made my life and work at USC an enjoyable experience.

Thanks to all friends for a valuable moment at USC starting with my roommates who provide the pleasant time in our apartment for years: P'nun Sunan Tugsinavisut, Barge Nopparit Intharasombat, Oh Ekaluck Chaiyaporn, Tong Kanate Ungkasrithongkul, P'a+ Suparerk Manitpornsut, Sia Suwicha Jirayuchjaroensak, Nok Pornthip Saelim, Ed Matus Saigamthong, Dej Peeradej Supmonchai and P'nui Siwas Chandhrasri. I had a challenging moment playing a hoop with Basketball gangster: Pop Phipat Phihakendr, P'wit Witaya Sungkarat, P'tu, Palmy Piyamaporn Kittimonthorn, Nancy Aphaphorn Kittimonthorn, Pun Thinnaphan Wanglee, Jay Nakarin Netcharussaeng and Matt Dr. Poonsuk Lohsoonthorn. Other Thai friends also make my life enjoyable on and off campus: P'ake Dr. Phunsak Thienvviboon, P'a+ Dr. Wuttipong Kumwilaisak, Amm Parichat Sakulphramana, Omm Ramanee Sakulphramana, P'sak Dr. Somsak Datthanasombat, P'Golf Dr. Piyapong Thanyasrisung, P'a+ Dr. Dhawat Pansatiankul, Prim Primrose Puempoon, Dle Nuchada Noochprayoon, Chompoo' Salida Sangsod, Mon (Sirimon Usap, Phet Phetrada Shenkrua, Pae Akkaya Shenkrua and their mom Benjamat Shenkrua, P'ake Boonake Champunod, Pang Pornsarun Wirojanagud, Ron Ranaphoom, Paew Warapa Bunnapasak, P'Oh Chanastha aeimketkeow, Yai Hongsuda Tangmunarunkit), Sethavid Gertphol, Job Chartchai Meesookho), P'a+ Dr. Dhawat Pansatiankul, Pin (Nantakan, P'nom Panom Intarussamee, Big Wibool Piyawattanametha, Somkiat Kraikriangsri, Earth Pattra

Chairojnitikorn, Sharon Dr. Shan Liu, Eddy, Dr. Alex Ossadtchi and many others. Without your presence, I would live my life without the enjoyable moment.

Finally, I would like to express my deepest appreciation to all following people who have been at my side giving me the tremendous support, constant encouragement, great guidance and the priceless love I needed from day one of my life in the US. In particular, I am extremely grateful for their supports, discussion, encouragement and inspiration during my difficult time. They taught me several of the most important lessons of my life: my father Phol Sagetong, my mother Siriporn Sagetong, my sister Orn Vichaya Sagetong, aunt Soon Srisoontorn Amornsiriwatanakul, aunt Duke Tossaphol Amornsiriwatanakul, aunt Toy+ Dujruethai Amornsiriwatanakul, Monk Jitre, and my friends Gail Gaywalee Yamskulna, Sunan Tugsinavisut, Noparit Intharasombat, Ekaluck Chaiyaporn, Kanate Ungkasrithongkul, Suwicha Jirayuchjaroensak, Pornthip Saelim, Primrose Puempoon, Phipat Phihakendr, Changsung Kim, Kemal Demirciler, Nguen and Janesakul Praemwat. At last, my love, support and faith go to Korn Kornwika Vongsariyavanich for sharing with me in the last period of my Ph.D. life.

Contents

List Of Tables

List Of Figures

Abstract

In this thesis, we address algorithms in the areas of image and video coding that aim at assuring reliable and secure communications. This work can be divided into 4 different parts: (i) two-dimensional dependent quantization, (ii) region of interest coding, (iii) multiple description coding, and (iv) watermarking.

The first contribution of the thesis is dedicated to the problem of two dimensional dependent quantization. Traditional greedy-search algorithms are popular due to their low complexity and low memory consumption, but they are far from optimal due to their dependency in coding. To resolve this problem, we develop an allocation technique, which can outperform greedy-search techniques. The proposed technique enables efficient selection of the quantization parameter for each coding unit such that the required memory size and computation time can be significantly reduced, as compared to the exhaustive search approach.

The second part of this thesis studies rate-distortion (RD) modeling for wavelet-based codecs and analytical bit allocation based on the proposed RD model. The proposed algorithm efficiently allocates different numbers of bits to different parts

of an image. This type of problem appears in application such as Region of interest (ROI) coding. The proposed scheme provides good estimates of the desired operating parameters with low complexity, e.g., without requiring an explicit generation of all possible RD operating points.

The third part of the thesis is devoted to the problem of error-resilient coding of images. We present Multiple Description Coding (MDC) as a data compression algorithm that provides efficient recovery from data losses for robust communication over erasure channels. The performance analysis and simulation results show that the proposed technique provides graceful degradation in the presence of channel erasures. In addition to good error resilience, our approach also provides simple adaptation to the changes in channel behavior by taking into account the proposed RD model during the optimization.

The last part focuses on video watermarking, and shows the development of a security mechanism for digital movie distribution. A known signal (a so-called watermark) which uniquely identifies the owner and authorized buyers is inserted into the copyrighted movie, in such a way as to be imperceptible. We determine where to cast and how to insert the watermark such that (i) the watermarked movie will be perceptually indistinguishable from the original one and (ii) the watermark will be present in the watermarked movie even after it has been attacked.

Chapter 1

Introduction

1.1 Motivation

Recent advances in computing and communication technology have stimulated research interest in the processing and distribution of multimedia information over wired and wireless networks. Data compression plays an important role in multimedia communications, since a compact representation of the large multimedia data sets leads to a more efficient usage of network bandwidth. Unlike traditional textual data, digital multimedia information (e.g., audio, speech, still and moving images) have three important characteristics. First, there exists a strong correlation between the samples of multimedia data. Second, multimedia information can tolerate losses, i.e., the signals can be reconstructed after decoding with various levels of quality. In other words, it is acceptable to have

a reconstruction that does not exactly match the original as long as the reconstruction does not have perceptually annoying artifacts. In this way, the lack of exact reconstruction may be acceptable, which results in much higher compression ratios. Third, there is a strict rule on the timing delay of decoding schedule but only in some cases, i.e., speech, audio and video. Clearly, text transfer is not subject to timing constraints since there is no schedule to decode the data. For example, sending an email requires lossless transmission since even one bit loss may alter the meaning of what was transmitted. Typically, text has to be lossless but is not subject to strict delay constraints.

In order to gain efficiency in coding and transmission of digital media content, the system must resort to some sort of compression such that the channel bandwidth or the storage space can be efficiently utilized. Rate-Distortion (RD) theory [11] provides a starting point by formalizing the tradeoff between coding bitrate and reconstruction error. Optimization in an RD sense has played a major role in recent image and video coding research [52, 57, 27]. Since sources such as images and video have a great deal of correlation from sample to sample, more efficient coding can be achieved by predicting samples based on previously transmitted samples. The difference between the sample value and the prediction will be encoded and transmitted. Clearly, it takes fewer bits to encode differences than it takes to encode the original sample as long as correlation exists. To

exploit this correlation between the coding units in multimedia data, these differential coding schemes have been used in many standards such as JPEG [53] for image compression and MPEG-1, 2, 4 [28, 29, 85] and H.26x [30, 31] for video compression.

With the expansion of digital image applications, not only is it important to provide good RD performance, but providing a functionality to process different portions of image with different fidelities is desirable as well. In recent years this non-uniform distribution of the image quality has been incorporated into image coders such as EZW [71], SPIHT [69] and JPEG2000 [6]. This feature is employed in applications that require images to be coded in such a way that the end user can view some portions of the image with higher decoding quality than the rest of the image. This form of progressive coding is called Region of Interest (ROI) coding. It enables the ROI to be reconstructed more quickly than the rest of the image, achieving a higher quality level as well.

In the presence of lossy channels, the extensive use of predictive coding by the source coder, in order to achieve better compression, makes the compressed data vulnerable. Thus considering only the compression mechanism without taking into account the transport mechanism may not be sufficient since failures in the transport could lead to very low quality reconstruction at the receiver. To gain robustness together with compression efficiency research has been very active in

the area of joint source channel coding (JSCC) [83]. The goal of these techniques is to provide error-resilient coding at encoder and/or error concealment at decoder such that the overall quality is maximized. In most of the current networking technologies such as Integrated services digital network (ISDN), Cellular networks, wireless LAN or Broadband wireless IP networks, packet losses can be frequent [81]. One main reason of the packet losses is network congestion, which continues to be commonplace nowadays. For instance, when the local buffer at a switching node or router suffers overflow, packets have to be dropped. For some applications, there are strict constraints on the end-to-end delay. Packets that arrive after their scheduled display time will be considered lost. In channels with high bit error rate or low signal to noise ratios, packets that contain detectable but uncorrectable bit-errors will also be considered lost.

Recent years have witnessed a proliferation of illegal copying of digital content, which is facilitated by the ease with which digital content can be disseminated over today's networks. Clear examples can be seen in the music and movie industries. Most songs currently available in the market can also be downloaded with good audio quality in the widely-used compact MP3 format. Cryptographic systems based on well-established algorithms (e.g., cryptosystem by Rivest Shamir Adleman (RSA) [36, 59] or Data Encryption Standard (DES) [21, 76]) are not sufficient to protect the data since their intent is to provide for a secured communication or data access control such that only the *authorized* viewers (the ones

who possess a proper key) can access to the scrambled signal. However once the data is descrambled to be viewed, these *authorized* customers might be the ones who then make a large number of copies and distribute them to others.

Research in the area of digital watermarking has primarily focused on ways to relieve this problem. In such methods, a known signal (watermark) is inserted into the copyrighted data. For the purpose of copyright protection, the casting watermark contains information about the owner and the authorized recipient. The watermark has to be robust to the deliberate or unintentional attacks. Once the copyrighted content is illegally distributed, the embedded watermark is extracted and used to place a claim of ownership and uniquely identify the source of the leak. In this way, appropriate action can be taken (punitive damages sought and security tightened, for instance).

In this thesis, four practical issues are addressed as follows. Bitrate-constrained allocation techniques for 2D dependent predictive coding and Region of interest coding are discussed in detail in sections 1.2 and 1.3, respectively. Novel channel-adaptive reliable transmission mechanisms that are robust in the face of the network impairments are addressed in Section 1.4. Secured distribution system to protect media against illegal attacks are presented in Section 1.5 and the chapter concludes with an overview of contributions of the thesis in Section 1.6.

1.2 Two-Dimensional Dependent Quantization

In recent years there has been a substantial amount of interest in *one-dimensional* (1D) predictive coding (or dependent quantization) problem [57, 27, 26]. To exploit the correlation appeared inside an image for image coding or within a frame in video coding, the introduction of *two-dimensional* (2D) differential coding has played a major role in reducing the amount of information needed to reproduce the signal at the decoder. The bit allocation problem then requires considering a 2D set of coding units (e.g., DCT blocks in standard MPEG-4 coding [85]), where the rate-distortion (RD) characteristics of each coding unit coding depend on one or more of the other coding units. As an example, in this thesis we consider MPEG-4 intra-coding [83, 85], where in order to further reduce the redundancy between coefficients both the DC and certain of the AC coefficients of each block are predicted from the corresponding coefficients in either the previous block in the same line (to the left) or the one above the current block. This 2D correlation leads to 2D dependency. Finding the optimal solution to this bit allocation problem may be a time-consuming problem, given that the RD characteristics of each block depend on those of the neighbors. One recently proposed technique to perform the bit allocation is to use the one-dimensional Viterbi algorithm (1D-VA) across rows via row-column iterations, with or without feedback from rows [40, 46, 34, 49]. However these approaches may not be efficient for problems with

2D dependencies, since they are not truly 2D in nature. Greedy-search algorithms are also popular due to their low complexity and low memory consumption, but they may be far from optimal due to the dependencies in the coding.

As an alternative, in Chapter 2, we introduce an iterative message-passing technique to solve 2D dependent bit allocation problems. This technique is based on (i) Soft-in/Soft-out (SISO) algorithms, which are first used in the context of Turbo codes [8], (ii) a grid model [79], and (iii) Lagrangian optimization techniques [73]. Unlike previous works that use the grid model for an *unconstrained* minimization problem [79], we propose to use the grid model to solve a 2D dependent bitrate-*constrained* allocation problem. The iterative message-passing technique attempts to achieve a globally optimal solution via a local-metric computation and message passing.

In this work, the key observation to solve this problem is that it is possible to iteratively compute the soft information of a current DCT block (intrinsic information) and pass the soft decision (extrinsic information) or message to other nearby DCT block(s). When the messages from every block arrive, the global cost function is eventually accounted for at every block. At this point, a hard decision can be made on the coding choice of each block. To guarantee that each block can get the information from every block via the least number of iterations, we define a schedule for the block activation.

1.3 Region of Interest Coding

Recently many researchers have focused their attention on coding schemes that can provide different level of fidelity within a given image. The part of the image that is reconstructed at higher quality is called Region Of Interest (ROI). With the limited bandwidth available on the Internet or over wireless networks, one might prefer to have higher decoding quality in some specific spatial portion of the image while maintaining acceptable quality in the rest of the image, so that the most important parts of the image can be displayed first, while the rest of the image is being downloaded. For higher bandwidth channels, the ROI concept may seem to be pointless. However, it is useful, for example, when the Internet user downloads image from a website that provides ROI coding. Since ROI is reconstructed faster than the rest of the image, the end user can decide to terminate the transmission as soon as they are satisfied with the reconstructed image. Due to the attractive characteristics of ROI coding, several research communities have attempted to incorporate this feature into their compression applications. For instance, in the signal compression community, the JPEG2000 standard [6] includes ROI functionality. Furthermore, in biomedical engineering application lossless ROI coding is desirable so that at some specific parts of the image of the patient body can be reproduced without losses [86, 54].

In Chapter 3, we address the problem of allocating bits to the different regions in an image coded with a progressive wavelet coder such as SPIHT (Set Partitioning in Hierarchical Trees) [69] or JPEG2000 [6]. This type of problem appears in ROI coding [3, 84]. The wavelet coefficients are divided by different factors before coding to enable different bit allocations to different regions, because the coefficients in each region are refined at different speeds. We call this dividing factor a **priority scaling factor** (psf).

For a given set of relevant distortion criteria, the best psf can be achieved by exhaustive search of all the possible psf values, after having explicitly measured the RD characteristics for each candidate psf values. A design based on empirical data could start by measuring overall RD operating points at a number of different psfs, and then proceed to select the best psf for a given criteria such as, for example, the relative quality between the ROI and the rest of the image. It is clear that an extensive empirical data generation process is required and the bit allocation is limited in that only a limited number of operating points can be chosen.

As an alternative, we introduce a novel Rate-Distortion (RD) model for images coded with a progressive wavelet coder and especially designed to capture RD behavior when different parts of an image are refined at different speeds. Our model is an extension of Mallat's model [42], which takes into account that the rates used are not necessarily the same throughout the image. Because of the

different rates, certain modeling approximations (e.g., those for coarse quantization) can not be used uniformly throughout the image. The main contribution of this work is thus to provide a novel analytical bit allocation technique based on the proposed RD model to determine what the psf should be, given criteria such as relative importance of the regions in ROI coding.

1.4 Error Resilience Coding

A common current scenario is that of transmission over a shared network such as the Internet, where there are no quality of service guarantees. Thus, multimedia data is packetized for transmission but, given that there are no priorities, any packet could be lost during transmission. Traditionally there have been two error control strategies mainly used so that reliable reproduction of data can be obtained: Forward error correction (FEC) [40] and Automatic repeat request (ARQ) [26]. FEC schemes operate by employing additional redundant bits to both detect and correct at the receiver errors incurred during a transmission. Unlike FEC approach, ARQ involves detecting error and requesting a retransmission of those packets that were either lost or received with errors. Retransmission of lost packets (ARQ) is undesirable (or even impossible) in some applications such as real-time video conference or low-bandwidth wireless communication. This is because in the scenarios where the channel error rate is high

the request for frequent retransmission will result in increasing the amounts of data transmitted, possibly contributing to additional congestion and delay. In some cases, retransmission is not possible, e.g., when there is no back channel. FEC schemes thus seem to be the best choice for delay constrained and/or no feed back applications. However FEC approaches suffer from the so called *cliff effect* once the bit error rate is higher than the detection/correction ability.

As the demand for image/video transmission has triggered the development of several techniques to provide error resilient distribution over channels subject to losses, it is desirable to have a coding scheme that can enable the receiver to reconstruct with acceptable quality while using only the coding units that were received, *without* having to request a retransmission. In Chapter 4, we will address one such approach based on Multiple Description Coding (MDC). In MDC, some redundancy is retained during source coding so that, after appropriate packetization, if packet losses occur it is possible to recover by exploiting the redundancy (statistical or deterministic) between what was received and what was lost. MDC has become popular for real time applications as it provides graceful degradation and does not require retransmission. In an MDC system, the signal is decomposed into several packets. If only some packets are correctly received then the decoder can reconstruct to an acceptable quality level. However, if all packets are received, information from one packet augments information from others, so that higher quality can be achieved relative to the case where only some packets

are received. This makes the system scalable as the quality improves when the packet loss rate is lower. Since MDC does not require prioritized transmission it can be used with current Internet protocols, such as UDP [75].

A fundamental design parameter in an MDC system is the level of redundancy. Higher redundancy provides more error protection, and therefore, ideally, one would like to match the redundancy to the channel characteristics, and in particular to be able to change the level of redundancy if the channel conditions are time varying. It is clear that such a trade-off exists, as the level of redundancy should increase when the packet loss rate increases, at the cost of some degradation in the corresponding error free performance. While MDC techniques have shown some promising results, one potential drawback is the fact that changing their redundancy level may entail significant changes to the system [82, 24, 70, 48, 55]. As an example, MDC techniques based on transform coding [82, 24] would require a modification of the transform at encoder and decoder each time the channel conditions change. Since the level of redundancy should be adjusted to match the specific channel conditions, the difficulty in adapting can be a significant problem for time varying transmission scenarios such as for real-time communications over wireless IP networks, due to the fluctuations of wireless channel conditions.

In Chapter 4, we propose a simple approach for MDC that involves using a polyphase transform and deterministic redundancy. Each sample of input data is transmitted several times, with different coding rates. This approach is useful

in that it greatly simplifies the design of an MDC scheme, since the rate allocation determines the amount of redundancy to be introduced in the signal that best matches a given target packet loss rate. Given that the decoder remains unchanged when the bit allocation changes it is possible to adapt very efficiently to the changes in channel behavior without requiring a change in the packet sizes, or the structure of the decoder. Also, this provides a great deal of flexibility as it enables the choice of redundancy to be almost arbitrary. The proposed RD model in Chapter 3 can be used to eliminate the need for RD data generation and determine the optimal level of redundancy for a given bitrate and packet loss rate.

1.5 Secured Distribution of Copyrighted Data

Besides the problems one may encounter in distributing content when packet losses occur, copyrighted content presents additional challenges, as the goal is to ensure that illegal copies of the content are difficult to make and that those who make them can be tracked down. In this work, we present a mechanism and algorithm for creating undeniable watermarks to provide a secured distribution. This allows the non-repudiation of watermarked content. More specifically, we study *where* in the source domain the watermark bits should be designated and *how* to cast the watermark bits into these locations. To satisfy the non-repudiation

watermark schema requirements, watermarking algorithms must have the following characteristics: First, the watermark should use a key from a large number space such that no two keys are likely to be identical if keys are chosen at random. Second, the watermark key can be detected given information other than the value of the key itself. Third, each copy has a unique watermark associated with a distinguished key for transaction information in digital content distribution systems. This key should be usable to detect the watermark. We consider a digital cinema scenario, where at least 56 bits of watermark payload are needed to identify a movie at each specific theater and for each show time. Fourth, it is nice to have blank detection or semi-blank detection watermark. The watermark detection agent should be able to detect a watermark with no or very limited information. Last and most importantly, we need to create non-fragile or robust watermarks. This is the most important requirement in this proposed watermark schema. To verify the watermark key, the watermark agent needs to reconstruct the encrypted key-stream exactly. This non-repudiation watermark schema can be used for copy source tracking in a secure digital content distribution system, which uses broadcasting technologies, such as satellite or multicast.

In the digital distribution system, we assume that the content owner or provider uses outside agents to distribute its content. Digital content, which is watermarked by distribution agents will be undeniably recognizable by the content provider as originating with that distribution agent. That is to say that

given certain distribution agents, the content provider will be able to tell, which distribution agent watermarked the content. The system does not allow a given distribution agent to watermark content so that it appears to have been watermarked by another agent. It also does not allow the content provider to watermark content that would appear to have been watermarked by a particular distribution agent. This allows the content provider to place a high degree of trust in the identification of the distribution agent and trace "leak" locations of pirated copies of videos.

In Chapter 5, we propose novel watermark embedding and detection methods to address the problem of copyright tracking of movies distributed to theaters by satellite. Temporal and spatial wavelet transformation as well as feature-based watermark embedding procedures are deployed such that the proposed watermarking algorithm is able to achieve a compromise tradeoff between visual quality and robustness. The multi-resolution nature of the wavelet transformations of a video sequence makes the watermark very robust and secure. Wavelet features make very large watermark patterns possible, which allows unique dynamic labeling of a large number of videos. Changing some of the features in video wavelet transformation domain according to the defined rules makes the watermark detection semi-oblivious, which allows watermark detection very robust and independent without the original sequence. Our proposed watermark can also be detected robustly after many kinds of malicious attacks.

1.6 Outline and Contributions

The main contributions of this research are

- *Message-passing algorithm for 2D dependent bit allocation.* In 2D predictive coding, intensive complexity and large storage requirement are required to perform exhaustive search. Greedy-search methods relax these requirements but still suffer from no guarantee of optimum solutions. In this work, we have proposed to use the message-passing algorithm based on a grid model to perform bit allocation for 2D dependent quantization problem. We show that with slightly higher computation and memory cost, we achieve a significant performance gain over the greedy-search method [66].

- *Analytical model-based bit allocation for region of interest coding.* The main novelty of our method is that it proposes an analytical bit allocation technique based on a novel RD model. This is used to encode different parts of the image at different bitrates by using a wavelet-based codec. The proposed algorithm provides an accurate estimate as compared with the empirical-based technique, and can achieve significantly reduced complexity [64, 16, 65, 17, 18].

- *Optimal bit allocation for channel-adaptive multiple description coding.* We have proposed a simple approach for MDC that uses a polyphase transform

and deterministic redundancy. This approach is useful in that it greatly simplifies the design of a MDC scheme, since the rate allocation determines the amount of redundancy. Moreover, it provides a great deal of flexibility as it enables the choice of redundancy to be almost arbitrary. We have introduced an optimal bit allocation algorithm based on the proposed RD model that allows us to select the amount of redundancy to be introduced in the signal that best matches a given target packet loss rate. Chapter 4 covers this part of our work [63, 64, 62].

- *Dynamic wavelet feature-based watermarking for copyright tracking in digital movie distribution system.* For a digital media content, other than providing a reliability against packet losses, the security is one important issue to be concerned. We consider the problem of where to put the watermark an how to cast the watermark into the selected location such that it is imperceptible and robust against the expected attacks. We have proposed non-repudiable watermarking schemes for copyright tracking of movies distributed to theaters by satellites. The proposed algorithm is robust against several types of attack with unnoticeable visual difference between the original movie and the watermarked one [67, 68, 89].

Chapter 2 is devoted to two-dimensional constrained dependent quantization, where the dependency comes from using a predictive encoding environment. The

analytical bit allocation technique for ROI coding is proposed in Chapter 3. This technique is based on a novel RD model to provide the efficient RD data generation. In Chapter 4 the channel-adaptive bit allocation for MDC is proposed. We present the way to achieve the reliable transmission in the networks subjects to losses. Our scheme provides an efficient method to adapt to the changes of channel conditions. Chapter 5 studies the problem of providing confidential distribution of copyrighted media content. A novel watermarking algorithm is proposed to cast the watermark in the wavelet domain by adjusting the features of the wavelet coefficients in such a way that it will be unnoticeable and robust against illegal attacks. Finally in Chapter 6 extensions and future work are described.

Chapter 2

Message Passing Algorithm for 2D Dependent Bit Allocation

In this chapter, we study efficient bit allocation algorithm under 2-dimensional dependent environment [1].

2.1 Introduction

The general dependent bit allocation problem can be found in many image/video coding applications such as spatially and temporally dependent coding in MPEG-1, MPEG-2 and MPEG-4 standards [85, 83, 52] and H.26x [30, 31]. In dependent coding one chooses quantizers, or number of bits for each coding unit (e.g. DCT-block or frame), but the actual rate and/or distortion depends on the neighboring coding units. That is, a set of available R-D operating points for some coding

[1] Work presented in this chapter was published in part in [66]

units depends on a particular choice of R-D operating points for other coding units. Clearly, although an exhaustive search method provides a global optimization, it is a time-intensive process and requires large memory storage, which makes it impractical. This is because in a dependent coding framework where there exists a prediction, the number of R-D points to be computed grows exponentially. A majority of research dealing with dependent-coding environments has focused on greedy search approaches, which performs a local optimization. Although greedy approaches relieve the complexity and storage requirements, they do not guarantee an optimum solution nor do they guarantee solutions within a certain range of the optimum. Therefore, it would be useful to apply an alternative searching scheme that can efficiently provide a nearly optimal choice while maintaining acceptable complexity and storage requirements.

The one-dimensional Viterbi Algorithm (1D-VA) [40] has been successfully used to reduce the complexity problem for one-dimensional (1D) dependent systems, but it can not be used directly for two-dimensional (2D) dependent systems as there is no single way of ordering 2D data. The lack of a natural order in two dimensions means that there is no way to map this problem into a 1D-VA optimization. Several researchers have attempted to shed light on the 2D dependent problem. They have initially focused on the application of the 1D-VA across rows and columns via row-column iterations of hard decisions, with or without decision feedback from rows, which is known as Decision-Feedback VA

(DF-VA) [46, 34, 49]. Based on an iterative decoding algorithm used for Turbo codes [8, 5], an iterative Soft-in/soft-out (SISO) algorithm has been proposed as a tool for digital image processing/compression problems, e.g., digital image halftoning [7] and near-lossless compression with row/column processing [9].

Although passing soft information typically provides a significant increase in performance, it is a complex process [79]. Complexity reduction techniques are thus needed, which may exploit special characteristics of the problem. For example, a grid model [80] has been introduced to solve the digital image halftoning problem. It is a graphical model containing nodes and edges connecting the nodes. Each node is connected to others by edges and has a support region representing the dependency to other nodes. The overlap information between two nodes represents the intersection between the support regions of two consecutive nodes. For example, in the image halftoning problem, if a 3×3 filter is used, a support region is a 9-pixel footprint of the filter centered at the pixel being considered. The overlap information is defined as the 6 pixels corresponding to the intersection between consecutive 9-pixel footprints of the filter. The soft information passed between two coding units (pixels) is the information corresponding to the 6 pixels that correspond to the intersection between consecutive 9 pixels footprints of the filter used in halftoning. Therefore, the soft information takes on $2^6 = 64$ possible configurations.

A standard message-passing algorithm is a method that utilizes a local message computation and distribution such that nearly optimal solutions will be achieved after a few iterations. In a grid-model-based message-passing algorithm, a node receives messages from other directly connected node(s) and, in turn, sends messages to every connected nodes. A message represents a measure of the quality of the possible configurations of the overlap information. The complexity of this algorithm can be kept low due to the local computation. In this work, based on the grid model, we apply an iterative message-passing technique, which attempts to achieve a globally optimal solution via iterative local-metric computation and message passing. Unlike the previous work where the grid model was used for digital image halftoning [80], here we use the grid model to solve the 2D bit allocation problem that arises in MPEG-4 intra-coding [85]. In MPEG-4, instead of performing only DPCM-coding to the DC coefficients in a horizontal direction as in a general block-based video coding standards [83], both DC and certain of its AC coefficients will be predicted from adjacent blocks in 2D direction. More specifically, either the block on the left or the block above the block being considered can be chosen as a predictor as shown in Figure 2.1. It is worth nothing that in H.264/AVC [33] a similar prediction is used in the spatial domain by referring to neighboring samples of already coded blocks, in contrast to MPEG-4 standard where the prediction is conducted in the transform domain.

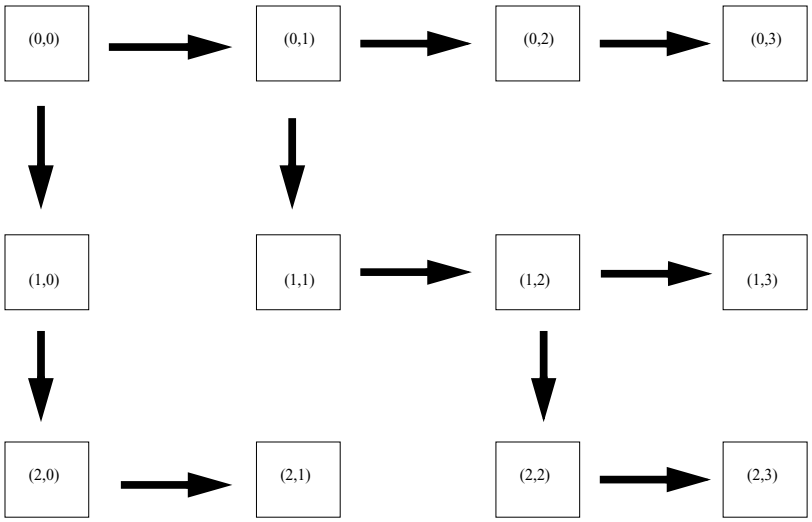


Figure 2.1: An example of the 2D dependent structure of MPEG-4 intra-frame of size 24×32. There are 12 blocks of size 8×8.

It is worth noting that in the digital image halftoning work [80] there exists only a *local* dependency that affects 8 neighbors around a given pixel.

Unlike the *local* dependency[2] that exists in digital image halftoning application, the RD characteristics of each DCT block in MPEG-4 intra-coding mode depend on the RD characteristics of *all* previously coded DCT blocks. Based on the original grid model [80], the overlap information is defined as the quantization parameters that affect the actual rate and distortion of the coding unit being considered. For example in Figure 2.1, the arrows show the direction of the dependency as specified by MPEG-4 texture-coding system. It implies that the RD curve of node (2,3) depends on the RD curves of nodes (2,2), (1,2),

[2]8 neighbors are affected around a given pixel. Note that extent of the dependency is a function of the filter that is chosen for the halftoning problem

(1,1), (0,1) and (0,0). Thus the overlap information is the set of the quantization parameters of nodes (1,2), (1,1), (0,1) and (0,0). Given the large number of possible configurations, using the original grid model and a standard message-passing algorithm [80] would not result in complexity reduction as compared to the exhaustive search method.

We propose here an approximation technique that modifies the grid model and the standard message-passing algorithm so that a solution can be achieved with significant reduction in complexity and memory consumption as compared with the exhaustive search method. Compared with the greedy search approach, the proposed method consumes slightly more memory. Although it requires higher complexity, it is polynomial-time solvable with respect to the input image size (the number of blocks). Additionally, the proposed technique provides significantly better performance. The main contribution of our work is to propose a message-passing algorithm to efficiently allocate bits to the block in an intra frame where prediction is used in between blocks, such that total distortion of the frame is minimized under the budget constraint.

This chapter is organized as follows: in Section 2.2, we introduce the iterative message-passing algorithm for the 2D dependent bit allocation problem arising in MPEG-4 coding. We explain this algorithm by starting with a definition of the *messages* or soft information that is exchanged between neighboring nodes. We then address how these messages will be utilized such that a near globally

optimal solution can be achieved. An analysis of the complexity and memory is provided in Section 2.3. It includes a comparison among the proposed message-passing technique, the exhaustive search method and the greedy search approach. In Section 2.4 we apply the proposed algorithm to solve a temporally dependent coding problem, which is a simplified version of 2-D dependent coding problem. Experimental results are shown in Section 2.5 and conclusions are provided in Section 2.6.

2.2 Iterative Message Passing Algorithm

The recent interest in SISO algorithms has led to a flurry of proposed approaches in various communications engineering applications where SISO techniques are used to provide a reliability measure on the possible signal choices [8, 5]. An example application is the efficient decoding of Turbo codes. The main idea of SISO algorithm is to iteratively produce soft information from the current subsystem (intrinsic information) and pass the messages (which become an incoming extrinsic information) to other nearby subsystem(s) as shown in Figure 2.2.

These subsystems need to communicate with each other because they share information. More specifically, at each activating subsystem, the accumulated information, i.e., the intrinsic information, is determined by combining information from the subsystem to be processed with the extrinsic information passed

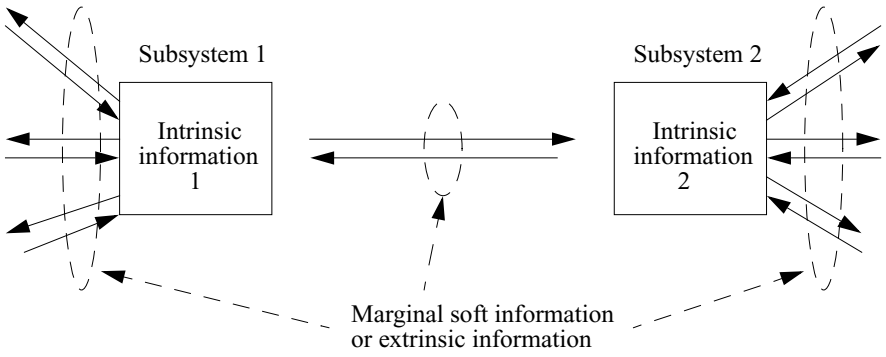


Figure 2.2: General message-passing between subsystems. The intrinsic information is computed by combining the received messages (the incoming extrinsic information) from other subsystems. The outgoing extrinsic information in each direction is computed by performing the optimization over the intrinsic information and then removing the direct effect of the incoming extrinsic information in that particular direction to avoid having the repeated version of the information in the system.

from other subsystems. Afterwards, the intrinsic information is used to generate an updated message (the outgoing extrinsic information) that is passed to other subsystems, after removing the direct effects of the incoming extrinsic information that was used to construct the intrinsic information. We define an iteration as the successive activation of all subsystems. Iterations proceed until a stopping condition is met. In general SISO algorithm used for Turbo codes, the extrinsic information and the intrinsic information are probabilistic (e.g. a posteriori probability (APP) estimates).

Recently SISO algorithms have been proposed as a tool for digital image processing/compression problems, e.g., digital image halftoning [7] and near-lossless

compression with row/column processing [9]. Although passing soft information typically provides a significant increase in performance, it is still a complex process [79], as mentioned in the previous section. A grid model [80] has been introduced to facilitate computations on a previously proposed structure [9] and improve its performance. In this work, unlike the previous work [80] where a grid model was used for a problem of *local* dependent coding, which arises in applications such as digital image halftoning, we modify the grid model to solve a problem of *global* dependent coding arising in MPEG-4 intra-coding [85]. This problem involves additional difficulties as there exist 2D global dependencies where the RD data of each coding unit depends on the RD data of *all* previous coding units (not only on the previously immediate one) and the coding units can be located anywhere in 2D space, as shown in Figure 2.1.

2.2.1 Problem Formalization

In a DCT-based compression scheme, the encoding rate and the associated distortion of a video frame are determined by how coarsely the DCT coefficients are quantized. The coarseness of the quantization can be scaled by adjusting a quantization parameter (i.e., quantization step size). The 2D dependent bit allocation problem for an MPEG coding system is therefore a problem of selecting the

encoding rate or the quantization parameter for each DCT block so as to minimize the overall distortion of the whole intra-frame, given that the encoding rates are restricted by a bitrate budget. More specifically, in MPEG-4 texture-coding when a given DCT block is to be intra-coded, some of its coefficients will be predicted from adjacent blocks in either vertical or horizontal directions, because in general there exists a strong correlation among these blocks in 2 directions. It is worth noting that the standard does not allow the block to be predicted from *both* directions. Only one predictor can be chosen. However, each block can be used as a predictor to either the block below or the one on its right or both of them. The choice of the most appropriate block (predictor) is made by measuring the picture gradient defined by the change of the DC coefficient. If prediction is in the vertical direction, the top row of coefficients is predicted from the block above so that only the difference between them needs to be coded. If horizontal prediction is chosen, the left column of coefficients is predicted from the block on the left so that again only the differences need to be coded. Two alternate scans are conducted depending on the prediction direction. It is worth noting that the choice of prediction is made before coding of the blocks, i.e., it does not depend on the quantization choice for each block.

Given the prediction chosen for each block, our goal is to determine the best quantization parameter (i.e., the right number of bits) for each block, $\mathbf{Q} = \{q_{0,0}, q_{0,1}, \ldots, q_{\frac{R}{8}-1,\frac{C}{8}-1}\}$ so that a given bitrate budget R_B is met. Here the

size of the image is $R \times C$ and $q_{i,j} \in \mathcal{Q}$ (finite set). Given that there are N admissible quantization parameters, $\mathcal{Q} = \{r_0, \ldots, r_{N-1}\}$ where r_i is the quantization parameter and $r_i < r_j$ if $i < j$. The goal is to minimize the total distortion in the intra-frame. Clearly, the right choice for the quantization parameter for a given block will depend on the quantization choice made at the block from which it is predicted. More specifically, the goal is to find a quantization parameter $q_{i,j}$ for a block $B_{i,j}$ so as to minimize an additive distortion with local dependency. Let $D_{i,j}(q_{i,j}|\mathbf{P}_{i,j})$ and $R_{i,j}(q_{i,j}|\mathbf{P}_{i,j})$, respectively, be the distortion and rate associated with block $B_{i,j}$, where $\mathbf{P}_{i,j}$ is a set of quantization parameters associated with all blocks that affect the RD curve of block $B_{i,j}$. For instance in Figure 2.1, $\mathbf{P}_{2,3} = \{q_{2,2}, q_{1,2}, q_{1,1}, q_{1,0}, q_{0,0}\}$. Finding the best quantization parameters for each of the blocks in the frame, $\mathbf{Q}^*$, can be stated as a constrained optimization problem, where each $q^*_{i,j}$ has to minimize the total distortion subject to a total bitrate budget constraint (the average bitrate per sample has to be equal to the budget, R_B); i.e.,

$$\mathbf{Q}^* = \arg\min_{\mathbf{Q}} \; D_T(\mathbf{Q}) \quad \textit{such that} \quad R_T(\mathbf{Q}) \leq R_B \tag{2.1}$$

where

$$D_T(\mathbf{Q}) = \sum_{i=0}^{\frac{R}{8}-1} \sum_{j=0}^{\frac{C}{8}-1} D_{i,j}(q_{i,j}|\mathbf{P}_{i,j}) \quad and \tag{2.2}$$

$$R_T(\mathbf{Q}) = \sum_{i=0}^{\frac{R}{8}-1} \sum_{j=0}^{\frac{C}{8}-1} R_{i,j}(q_{i,j}|\mathbf{P}_{i,j}). \tag{2.3}$$

2.2.2 Message-passing-based Bit Allocation Algorithm

Lagrangian optimization techniques [20, 88, 73] can be used to solve this problem by introducing a cost function, $J_{i,j}(\lambda, q_{i,j}|\mathbf{P}_{i,j}) = D_{i,j}(q_{i,j}|\mathbf{P}_{i,j}) + \lambda R_{i,j}(q_{i,j}|\mathbf{P}_{i,j})$, where a Lagrange multiplier, $\lambda \geq 0$, is used to trade-off rate and distortion. This leads to an unconstrained minimization of the cost function for the correct value of λ, namely λ^*, that is the one for which the quantizer allocation meets the budget constraint R_B. The best solution for a given λ^* is:

$$\mathbf{Q}^*(\lambda^*) = \arg\min_{\mathbf{Q}} J(\lambda^*, \mathbf{Q}) \tag{2.4}$$

where

$$J(\lambda^*, \mathbf{Q}) = \sum_{i=0}^{\frac{R}{8}-1} \sum_{j=0}^{\frac{C}{8}-1} J_{i,j}(\lambda^*, q_{i,j}|\mathbf{P}_{i,j}). \tag{2.5}$$

In an *independent* coding environment where each block's rate and distortion can be determined without knowing the quantization assigned to any other block,

the above minimization would be done *individually* for each block. In other words, $D_{i,j}(q_{i,j}|\mathbf{P}_{i,j}) = D_{i,j}(q_{i,j})$ and $R_{i,j}(q_{i,j}|\mathbf{P}_{i,j}) = R_{i,j}(q_{i,j})$. However, it is obvious that in a *dependent* coding environment, in order to quantize a given block the quantized version of its predictor blocks need to be known and the optimization thus becomes more complex. Moreover not only does the previous immediate quantized predictor block have to be known, but in effect all previously quantized blocks need to be known. That is, the distortion of the block being considered depends on the distortions of *all* previously coded blocks. In this work, our goal is to combine the message-passing technique with the Lagrangian optimization method to determine $\mathbf{Q}^*(\lambda)$ and iteratively change λ using the bisection algorithm until we find the best multiplier λ^* such that the total bitrate used is $R_T(\mathbf{Q}) = R_B$.

For the sake of simplicity, we will explain the message-passing algorithm for the 2D dependent bit allocation problem by using a concrete example as illustrated in Figure 2.1. The predictors of each block are pre-determined as described before. Given that each block can not be predicted from *both* the block above and the block on its left, the underlying graphical model contains no cycles, as shown in Figure 2.1.

Each edge, labeled by $e[(i,j),(m,n)]$, represents the edge variable between node (i,j) and node (m,n), i.e., $e[(i,j),(m,n)] = \{q_{i,j}, q_{m,n}\}$, where $q_{i,j}, q_{m,n} \in \mathcal{Q}$. It represents 2 quantization parameter variables corresponding to all possible

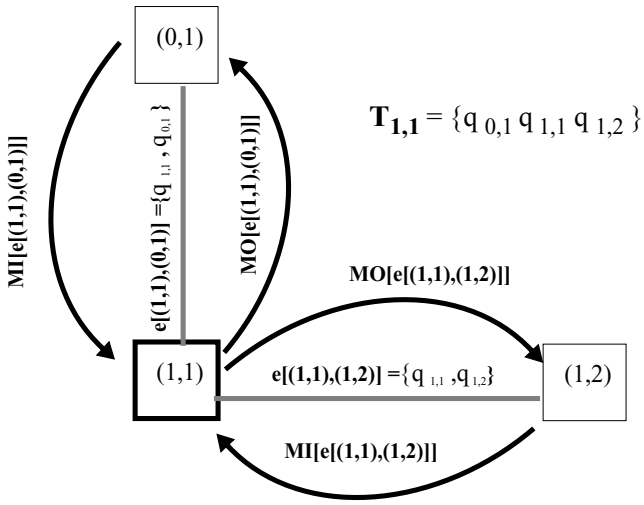


Figure 2.3: Illustration of edge variables $e[(1,1),(0,1)]$ and $e[(1,1),(1,2)]$, local dependency configuration, $T_{1,1}$, of node $(1,1)$, incoming mutual information $MI[e[(1,1),(0,1)]]$ and $MI[e[(1,1),(1,2)]]$ and outgoing mutual information $MO[e[(1,1),(0,1)]]$ and $MO[e[(1,1),(1,2)]]$ when node $(1,1)$ is considered

quantizer combinations of $q_{i,j}$ and $q_{m,n}$. Thus each edge takes on N^2 possible configurations. Each node will accept the messages or mutual data from other connecting nodes. At node (i,j), $MI[e[(i,j),(m,n)]]$ denotes the incoming message from node (m,n). A message represents an accumulated cost providing an updated knowledge about the cost function for each possible edge configuration. For each configuration the best decision of all quantization parameters and the cost associated with the best decision of all quantization parameters in that particular direction represents the message. $T_{i,j}$ denotes the local dependency configuration (variables) for node (i,j). It represents the quantization parameter

variables of all possible nodes directly connected to node (i, j) including itself, node (i, j).

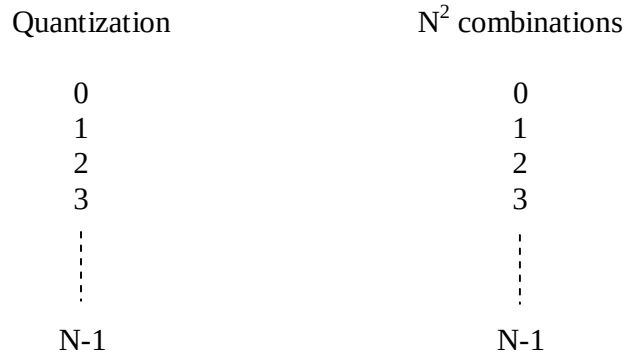


Figure 2.4: Illustration of maximum number of configuration, N^2

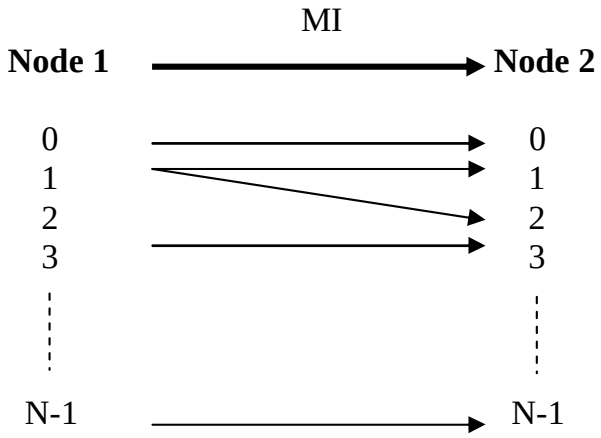


Figure 2.5: The information is passed from Node 1 to Node 2 based on the best decision for each configuration

For example in Figures 2.1 and 2.3, at node $(1, 1)$, variables corresponding to a $T_{1,1}$ are $\{q_{0,1}, q_{1,1}, q_{1,2}\}$. The edge between node $(0, 1)$ and node $(1, 1)$ is labeled by $e[(1, 1), (0, 1)] = \{q_{1,1}, q_{0,1}\}$. The message sent on the edge between node

$(1,1)$ and node $(0,1)$ can be represented as $MI[e[(1,1),(0,1)]] = MI[q_{1,1}, q_{0,1}] = J(q_{1,1}, q_{0,1})$ where $J(q_{1,1}, q_{0,1})$ denotes the accumulated cost sent from node $(0,1)$ to node $(1,1)$ for a given edge configuration. The edge and the message each contain N^2 possible configurations since there are 2 quantization parameter variables, i.e., $q_{1,1}$ and $q_{0,1}$. Figures 2.4 and 2.5 show the maximum number of configurations and the information passed between nodes. The final best quantization parameter will be chosen only at the last iteration by performing a hard decision. Similarly, the message over the edge between node $(1,1)$ and node $(1,2)$ is $MI[e[(1,1),(1,2)]] = MI[q_{1,1}, q_{1,2}] = J(q_{1,1}, q_{1,2}))$.

The intrinsic information, denoted by $INT[T_{i,j}]$ of node (i,j), is an accumulated cost of each block for each possible configuration. It can be computed as shown below by *combining* the best decisions on each configuration received from the incoming messages from the nodes directly connected to the node being considered.

$$INT[T_{i,j}] = \sum_{(m,n)\ :\ e[(i,j),(m,n)] \in E_{i,j}} MI[e[(i,j),(m,n)]] \tag{2.6}$$

This intrinsic information $INT[T_{i,j}]$ is determined from all incoming messages for a given configuration, $MI[(i,j),(m,n)]$, where node (m,n) denotes a set of all possible "nodes" *directly* connected to node (i,j). Note that $E_{i,j}$ is a set of all possible "edges" *directly* connected to node (i,j).

Note that each of $INT[T_{i,j}]$ is determined uniquely by the configuration $T_{i,j}$. At node $(1,1)$, the total number of possible configurations for $INT[T_{1,1}]$ is N^3 since there are 3 quantization parameter variables, i.e., $q_{1,1}, q_{0,1}$ and $q_{1,2}$, and each quantization parameter variable can be one of N admissible quantization parameter in $\mathcal{Q}$. Each configuration contains a list of the best decisions from other nodes corresponding to each possible choices of $q_{1,1}, q_{0,1}$ and $q_{1,2}$.

After the node has finished the combining process by accepting the messages from its neighboring nodes, we then need to return an updated message to each of those nodes. We call this process *the marginalization process*. The new outgoing extrinsic information returned on each edge can be determined as shown below.

$$MO[e[(i,j),(m,n)]] = MO[q_{i,j}, q_{m,n}] = \arg \min_{T_{i,j}:e[(i,j),(m,n)]} INT[T_{i,j}] \ominus MI[q_{i,j}, q_{m,n}] \tag{2.7}$$

where at node (i,j), $MO[e[(i,j),(m,n)]]$ denotes the outgoing extrinsic information from node (i,j) to node (m,n). The minimization operation is conducted for all consistent configurations corresponding to each edge variable with a particular value. The goal of the optimization is to determine the set of configurations that minimize the cost $INT[T_{i,j}]$. Finally the subtraction is performed by removing the direct effects of the corresponding old (or previously incoming) extrinsic

information, $MI[e[(i,j),(m,n)]]$, which was used to construct the intrinsic information, $INT[T_{i,j}]$. We show below how to extract extrinsic information at node (i,j) where $\ominus$ denotes the *remove* operation.

An example of how to compute the outgoing extrinsic information at node (1,1) is shown below. It is worth noting that the $MO[e[(1,1),(0,1)]]$ of node $(1,1)$ is the $MI[e[(0,1),(1,1)]]$ of node $(0,1)$.

$$\begin{aligned} MO[e[(1,1),(0,1)]] &= MO[q_{1,1}, q_{0,1}] \\ &= \arg \min_{T_{1,1}:e[(1,1),(0,1)]} INT[T_{1,1}] \ominus MI[q_{1,1}, q_{0,1}] \end{aligned} \quad (2.8)$$

$$\begin{aligned} MO[e[(1,1),(1,2)]] &= MO[q_{1,1}, q_{1,2}] \\ &= \arg \min_{T_{1,1}:e[(1,1),(1,2)]} INT[T_{1,1}] \ominus MI[q_{1,1}, q_{1,2}] \end{aligned} \quad (2.9)$$

To guarantee that each node receives the information from *all* other nodes via the least number of iterations, we define a schedule of the node activation. For the purpose of activation scheduling, we re-arrange the dependency structure from the original structure as shown in Figure 2.1 to a tree as shown in Figure 2.6. Since the graphical grid model is cycle-free, node $(0,0)$ is considered as the tree

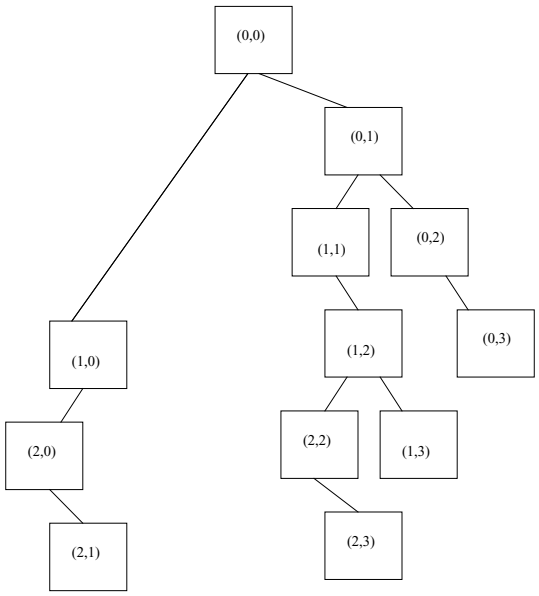


Figure 2.6: A dependency tree associated with a system shown in Figure 2.1

root and those nodes that are not used to predict any other node are considered leaf nodes (e.g., the nodes at positions $\{(0,3),(1,3),(2,1)$, and $(2,3)\}$).

We propose to start activating the nodes in the *downward* direction. That is activating the root node first and proceeding the activation following the tree path until all the leaf nodes are activated. Then we activate the leaf nodes and proceed upward to the root. It is worth noting that, unlike the downward direction activation, each intermediate node has to have the information from all other nodes below it. For example, the nodes at positions $(1,3),(2,3)$, and $(2,2)$ have to be activated before activating node at position $(1,2)$. Next we activate the node by following the trace we activated reversely. From Figure 2.6 as an example of upward direction, nodes should be scheduled as follows: (i)

$\{(0,0)\}$, (ii) $\{(0,1)\}$, (iii) $\{(1,1)\}$, (iv) $\{(1,0),(1,2)\}$, (v) $\{(0,2),(2,0),(2,2,)\}$ and (vi) $\{(0,3),(1,3),(2,1),(2,3)\}$. For a downward direction, the schedule is: (i) $\{(0,3),(1,3),(2,1),(2,3)\}$, (ii) $\{(0,2),(2,0),(2,2,)\}$, (iii) $\{(1,0),(1,2)\}$, (iv) $\{(1,1)\}$, (v) $\{(0,1)\}$ and (vi) $\{(0,0)\}$. This is the two-dimensional dependent bit allocation using message-passing algorithm.

Algorithm: 2D Dependent Bit Allocation using Message-Passing Algorithm

1. Start by setting $\mathbf{Q}_U = \{r_{N-1}, \ldots, r_{N-1}\}$ and $\mathbf{Q}_L = \{r_0, \ldots, r_0\}$.
2. Compute $\lambda = -\frac{D_T(\mathbf{Q}_U) - D_T(\mathbf{Q}_L)}{R_T(\mathbf{Q}_U) - R_T(\mathbf{Q}_L)}$.
3. Initialize all incoming extrinsic information, i.e., $MI[e[(i,j),(m,n)]] = 0$
4. Perform the message-passing algorithm.
   (a) Consider nodes in the order they appear in the scheduling list .
   (b) Combining process: determine the intrinsic information, $INT[T_{i,j}]$ as shown in Equation (2.6).
   (c) Marginalization process: determine the outgoing extrinsic information $MO[e[(i,j),(m,n)]]$ in each edge as shown in Equation (2.7).
   (d) Repeat step 4a at the next node in the scheduling list until we complete all the nodes in the list.

(e) When we finish the iterations, we compute the final hard decision by

$$\bar{q}_{i,j} = \arg \min_{q_{i,j} \in \mathcal{Q}} INT[T_{i,j}] \tag{2.10}$$

where $\mathcal{Q} = \{r_0, \ldots, r_{N-1}\}$.

5. If $R_T(\mathbf{Q}) = \mathbf{R_B}$ where $\mathbf{Q} = \{\bar{q}_{0,0}, \bar{q}_{0,1}, \ldots, \bar{q}_{\frac{R}{8}-1,\frac{C}{8}-1}\}$, $\mathbf{Q}^* = \mathbf{Q}$ and stop the algorithm. Else if $R_T(\mathbf{Q}) \geq R_B$, set $\mathbf{Q}_L = \mathbf{Q}$, or else if $R_T(\mathbf{Q}) < R_B$, set $\mathbf{Q}_U = \mathbf{Q}$. Go to step 2.

By repeating the activation of these nodes, we can pass the accumulated cost of each node to other neighboring nodes so that the overall optimal solution can be obtained since all dependencies are already taken into account. Note that the way we pass the best decision of quantization parameters from the nodes outside the window of local configuration may be similar to the greedy search approach. However only the best choices for the quantization parameter from the *previous* blocks are used in the greedy search approach. In the proposed technique, the best decisions from *both* previous and future directions are taken in account. Furthermore we iterate the node activation in the downward and upward directions and then make a final decision at the last iteration. Thus the performance of the greedy search approach can be considered as a lower bound performance of our proposed scheme.

2.3 Complexity and Storage Analysis

The exhaustive search method clearly provides the globally optimal solution while requiring a larger storage requirement and a very complex search. Although the greedy search approach achieves a significant complexity reduction and has low memory requirements, the solution obtained may not be a globally optimal solution. It is clear that the coding gain comes at the price of significant complexity. In this section, we will address the complexity and storage requirements of the proposed message-passing technique in number of operations and bytes, respectively, for the 2D dependent bit allocation framework we are studying. We also compare the results from the proposed algorithm with the exhaustive search method and the greedy search approach.

2.3.1 Complexity Analysis

To determine the overall complexity of the message-passing technique, we simplify the analysis by separating the computation into the number of operations used to perform (i) the data generation process (or RD population phase) and (ii) the searching process. The unit of operations for each is different. That is, one operation for the data generation process is the complexity used to compute the distortion and rate of one intra-frame. In other words, it is the number of times the encoder is called to determine the distortion and rate on one block. One

operation for the searching process is the complexity used to perform a numerical operation, e.g., plus, minus, minimization.

2.3.1.1 Complexity from RD population phase

For the proposed message-passing algorithm, the RD population phase is needed to determine $INT[T_{i,j}]$. Given N admissible quantizers and a maximum of 3 edges for each node, there are at most N^4 possible configurations of $INT[T_{i,j}]$. With an image size of $R \times C$, there are $S = \frac{R \times C}{8 \times 8}$ nodes. For L iterations in one intra-frame, we need LSN^4 or $\mathcal{O}\ (SN^4)$ operations.

The greedy search approach performs only a local search from the root to the leaves. The RD data for N choices of quantization parameters is made for each block separately and will not be changed in the future. Therefore we need SN^2 or $\mathcal{O}\ (SN^2)$ operations. Using the exhaustive search method, with all N^S possible configurations of quantization parameters, the complexity required to construct the RD operation points is $\mathcal{O}\ (N^S)$.

2.3.1.2 Complexity from searching phase

The intrinsic information, $INT[T_{i,j}]$ of node (i, j), has to be computed at each node. Given a maximum of 3 edges connected for each node, 3 addition-operations are required for each configuration in $INT[T_{i,j}]$ as shown in Equation (2.6). Given N admissible quantizers, there are N^4 configurations in $INT[T_{i,j}]$. Therefore,

$3N^4$ additive-operations will be needed to compute $INT[T_{i,j}]$ of node (i, j). To compute the outgoing extrinsic information, we need to perform N^2 comparison-operations for each configuration of each $MO[e[(i, j), (m, n)]]$. Given N^2 configurations on each edge and a maximum of 3 edges for each node, we need totally $3N^4$ operations. Therefore, the total number of operations used for node (i, j) is $3N^4$. For L iterations, we need $3LSN^4$ or $\mathcal{O}\ (SN^4)$ operations.

For the greedy search approach, only local minimization at each node has to be performed, i.e., N^2 comparison-operations. Thus it requires $\mathcal{O}\ (SN^2)$ operations to operate on all nodes. For the exhaustive search method, with all N^S possible configurations, to determine which choice yields the minimization, we need to go over all possible choices. This requires $\mathcal{O}\ (N^S)$ comparison-operations. Table 2.1 shows the comparison of complexity in number of operations among the described methods.

It is clear that although the message-passing technique yields higher complexity than the greedy search approach, it is polynomial-time solvable with respect to input image size (the number of blocks) as compared to the exponential-time solvable exhaustive search method. To summarize, the proposed algorithm offers an alternative way to provide the optimal solution (as can be provided by the time-consuming exhaustive search method) with a cost of an increase in complexity consumption compared to the greedy search approach. It however delivers a significant decrease in complexity compared to the exhaustive search method.

O(.)	Exhaustive	Greedy	Message-passing
Complexity (data generation)	N^S	SN^2	SN^4
Complexity (searching)	N^S	SN^2	SN^4
Storage (bytes)	SN^S	S	SN^2

Table 2.1: Comparison of the complexity in the number of operations and bytes among (i) the exhaustive search method, (ii) the greedy search approach, and (iii) the proposed message-passing technique. The complexity of the data generation phase is the number of times the encoder is called to code one intra-frame while the complexity of the searching phase is the number of numerical operations (+,-,Min) used to perform the searching process including the final hard decision process.

2.3.2 Memory Analysis

With the message-passing technique, only $MO[e[(i,j),(m,n)]]$, outgoing extrinsic information, needs to be stored since it will be used as an $MI[e[(i,j),(m,n)]]$, an incoming extrinsic information, to other neighboring nodes. Given N^2 configurations for each extrinsic information and a maximum of 3 edges for each node, there are $3N^2$ configurations for each node. With 1 byte to keep the quantization parameter index, we need a maximum of S bytes for each configuration to store the information of the quantization parameters. Therefore, we need $\mathcal{O}\ (SN^2)$ bytes.

Using the exhaustive search method, it requires to store all N^S possible configurations. Since each configuration stores a table of quantization parameter for each block, it requires S bytes. Therefore, the exhaustive search method consumes $\mathcal{O}\ (SN^S)$ bytes. Since the greedy search approach keeps only the decision of the quantization parameter of the block previously determined, it requires only

a table of the (locally) best quantization parameter to fill in, $\mathcal{O}(S)$. Table 2.1 compares the memory consumption in bytes among the searching techniques described. Clearly, the proposed algorithm requires significantly less memory than the exhaustive search method. It is worth noting that message-passing algorithm pay the price to achieve the better solution by consuming more memory than the greedy search approach.

2.4 Bit Allocation for Temporally Dependent Coding using Message Passing Algorithm

We now address the general *temporal* dependency quantization problem of which MPEG-x [28, 29] and H.26x [30, 31] are examples. The problem is to choose the quantization parameter for each frame such that the total distortion is minimized subject to a total bitrate budget constraint. It is worth noting that the *temporal* dependency quantization problem is a particular case of the *spatial* dependency quantization problem, in the sense that the structure is one dimensional and a one dimensional dependency may also occur in the 2D case. In video coding framework, several researchers have experienced the strong dependency in temporal domain [56, 52]. One example of this scenario is illustrated by Figure 2.7 for 4 frames (1-2-3-4). Each frame can be quantized using different quantization parameters where frame 2 is predicted from frame 1 and frame 3 is predicted

from both frame 1 and frame 2 ($I - P - B - I$). Clearly the set of available RD operating points for frame 2 (3) depends on a particular choice of RD operating point for frame 1 (2 and 4).

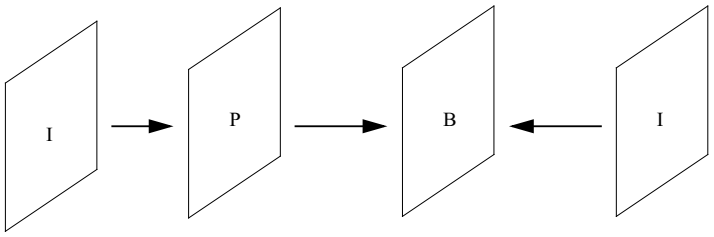


Figure 2.7: Example of temporal dependency for 4 frames (I-P-B-I)

Instead of using blocks as a coding unit as we have described for intra-frame coding, we consider one frame as the coding unit or the node. The goal is to determine the best quantization parameter for each frame such that the total distortion is minimized and the total bitrate is constrained under the given bitrate budget. The schedule of node activation starts from, for example, frame 1 toward frame 4 and then come back to the frame 1 to finish one iteration.

2.5 Experimental Results and Discussion

In order to confirm the validity of the proposed algorithm, for 2D dependent bit allocation for intra-frames in MPEG-4, we performed the experiment on the intra-frame (144×176 QCIF-formatted) video sequence Foreman and the still 512×512 gray-level images, Lena, Baboon and Plane. Note that still images

Image	Bitrate budget (bps)	PSNR (dB)	
		Greedy search	Message-passing
Foreman	0.80	27.00	27.67
	1.00	28.36	28.86
	1.20	29.37	29.72
Lena	1.70	25.87	26.32
	2.00	26.84	27.36
	2.30	27.60	27.82
Baboon	0.90	24.95	25.60
	1.00	25.59	26.34
	1.20	26.94	27.09
Plane	0.30	32.19	33.83
	0.50	36.17	36.65
	0.60	36.54	36.92

Table 2.2: Performance comparison between the greedy search approach and the proposed message-passing technique for 2-D spatial dependent coding problem

are considered here as intra-frames in a video sequence. We performed the experiment at different target bitrate budgets. As an image coder, we modified the normal JPEG coder to enable the MPEG-4 texture-coding feature. We first determined the quantizer for each block from the greedy search approach and our message-passing algorithm. The selected choice of quantizers was used as an input argument to the texture coding system. Experimental studies reveal that the system that uses the proposed message-passing algorithm is able to provide an achievable performance better than the greedy search approach by 0.57 dB on average as shown in Table 2.2.

Image	Bitrate budget (bps)	PSNR (dB)	
		Greedy search	Message-passing
Bream	0.70	50.00	50.97
	0.90	50.44	51.38
Coast guard	2.00	43.82	44.02
	3.00	46.14	46.83
Foreman	2.00	43.85	44.12
	3.00	46.61	46.83
Silent	0.70	42.98	44.85
	1.00	47.82	47.97

Table 2.3: Performance comparison between the greedy search approach and the proposed message-passing technique for temporally dependent coding problem

Similarly we validate the proposed scheme for the temporal dependency quantization problem. We perform the experiment for 4 different 10-frames QCIF-formatted video sequences of Bream, Coast guard, Foremen and Silent. The first frame is the I-frame and the rest are P-frame. The same admissible set of quantization parameter, $\mathcal{Q} \in \{2, 3, 4, 5, 6\}$, is used for each sequence. We performed the experiment at different target bitrate budgets. In Table 2.3 the proposed scheme shows promising results with improvements over the greedy search approach.

2.6 Conclusions

In this work we propose a new bit allocation algorithm for the 2D dependent problem that appears in MPEG-4 texture-coding system when intra-coding. Message-passing algorithm is applied in 2D dependent bit allocation problem. The proposed algorithm achieves promising performance while maintaining relatively low

complexity and storage requirements. We show that message passing technique can provide an achievable performance better than greedy search approaches with lower complexity and less storage requirement than full-search method.

Chapter 3

Analytical Model-based Bit Allocation for Region of Interest Coding

In the previous chapter we introduced a message-passing-based algorithm to solve a 2D dependent quantization problem. In this chapter, we consider another problem that arises when different parts of an image are required to be coded at different fidelities and speeds. This so-called region of interest (ROI) coding application is relevant to several communities, e.g., signal/image processing [6] and biomedical engineering [86, 54]. Our main motivation in this work is to achieve a simple algorithm to perform bit allocation in a ROI coding framework [1].

[1] Work presented in this chapter was published in part in [64, 16, 65, 17, 18]

3.1 Introduction

The bitplane-by-bitplane successive refinement used in most progressive wavelet coders, such as EZW [71], SPIHT [69] or JPEG 2000 [6], leads to a very simple technique to provide ROI coding [3, 50, 44, 14, 84, 13]. Roughly speaking, a progressive coder transmits information about large-magnitude coefficients before it transmits information corresponding to smaller coefficients. The goal in ROI coding is to transmit the region of interest with higher quality than other areas in the image. Since large coefficients are sent first, it is enough to divide the wavelet coefficients in areas outside the ROI by a factor greater than one so that they are transmitted later (on average) in the resulting bitstream. This dividing factor is called the *priority scaling factor* (psf). At the decoder, the reconstructed coefficients are then multiplied back by the corresponding psf before the inverse wavelet transform is performed. Since all the coefficients (after being divided) have been refined to a particular bitplane, it follows that those coefficients to which a $psf > 1$ has been applied will be more coarsely quantized, i.e., their binary representation will be "shifted" with respect to coefficients with $psf = 1$. Therefore on average more bits per pixel are used for the ROI than for the rest of the image[2]. This technique is known as a *divide-and-multiply* or *up-shift* method

[2]This motivates the need for a Rate-Distortion (RD) model covering all rates: if all coefficients are refined in parallel, it will be necessary to reach high rate in ROI, in order to obtain a reasonable (but lower) quality in the rest of the image.

and has appeared in Part 1 of JPEG2000 standard [6]. This technique is also useful for bit allocation inside each description in an MDC framework [63, 64], where we use the divide-and-multiply method to vary the quality of different parts of an image. Several versions of the same image are then transmitted, with each part of the image transmitted at various quality levels across images. More detailed descriptions will be provided in Chapter 4.

Prior works on bit allocation for ROI coding [3, 50, 44, 14, 84, 13] were based on heuristic techniques or required that rate and distortion characteristics be measured at each of the potential operating points. For example, a design based on empirical data could start by measuring overall image RD data at a number of different psf values, and then proceed to select the optimum or the most appropriate psf for a given criterion. There are two major drawbacks of this empirical-based method. First, it is obvious that the technique is limited in that the solution will have to be one of the discrete operating points that were explicitly measured, so that optimality may suffer if a bad choice was made of those discrete quantizers. Second, collecting the RD data may be complex. The more admissible quantizers are used, the more time will be needed to complete the RD data generation process. Although Part 2 of JPEG2000 standard [6] has adopted an ROI coding mechanism to analytically determine the dividing factor (Maxshift method [13, 78]), it is most appropriate for very high-quality ROI coding. This approach in JPEG2002 Part 2 is based on determining a

dividing factor that is sufficiently large to make the largest background coefficients smaller than the smallest non-zero ROI coefficients. In this way, all ROI wavelet coefficients will be transmitted first. Thus, the relative rates of the regions can not be chosen and after ROI refinement is completed we will have a very high rate ROI. This is equivalent to providing only a single operating point and limits the flexibility of the relative importance between ROI and the rest of the image. Therefore it would be useful to design a model-based bit allocation algorithm that can both efficiently work on arbitrary input images and enable fine granularity in the selection of RD operating points for the ROI and the rest of the image.

Several RD models for images have been studied with either high bitrate [11] or low bitrate [42] assumptions. In ROI coding, some parts of an image will be transmitted with higher quality or higher bitrate than the rest of the image. Thus we are motivated to define a model that will give accurate estimates and will provide good RD approximation for all bitrates. Rather than using separate models for different rate conditions we use a single model that can be valid in general at all rates. Thus, in this chapter, we modify Mallat's model [42], which performs accurately at low bitrates, to improve its performance at moderate and high bitrates.

The main contributions of our work are to (i) introduce a closed-form RD model for wavelet-coded images that can operate at any bitrate and (ii) use it within an analytical bit allocation technique for ROI coding to determine different

psf values such that a total rate budget is met and a criterion based on the relative distortions of the regions is optimized. We take SPIHT as an example for which our analysis is valid. Other progressive wavelet coders can be similarly modeled.

This chapter is organized as follows: in Section 3.2, we introduce the RD models, including models for each of the regions in an image when an ROI coding framework is used. We demonstrate the accuracy of the proposed RD model by comparing the results obtained by the proposed model with those obtained by an empirical search technique. Furthermore, we show the improvement by comparing the results obtained by the proposed model with those obtained by Mallat's model. We then present the analytical bit allocation based on the proposed model for ROI coding and analyze its complexity in Sections 3.3 and 3.4 respectively. The experimental results are shown in Section 3.5. The conclusions of this work are addressed in Section 3.6.

3.2 Analytical Rate-distortion Model

First, we briefly discuss the RD model for SPIHT originally proposed by Mallat and Falzon [42]. Given a total bitrate budget for an image B, the average quantization error $D(C_b)$ as shown below is the summation of the quantization error due to quantizing the significant coefficients E^{sig} and that due to setting to

zero the insignificant coefficients E^{insig}, divided by the total number of wavelet-transformed coefficients, N:

$$D(C_b) = \frac{E^{sig} + E^{insig}}{N} = \frac{\frac{C_b \Delta^2}{12} + \sum_{i=C_b}^{N-1} |w(i)|^2}{N}, \tag{3.1}$$

where C_b is the number of significant coefficients, that is, those with amplitude larger than Δ, the size of final quantization bin, i.e., the final threshold. Also, C_b is directly related to the average bitrate B via $C_b = \frac{NB}{6.6}$ which is obtained empirically. All N wavelet coefficients are sorted in monotonically descending order of magnitude, to obtain a list $\{w(i)\}$, for which $i = \{1, 2, \ldots, N\}$. E^{insig} is the energy of the last $N - C_b$ smaller amplitude coefficients since it is the error when all insignificant coefficients are quantized to zero, i.e., $E^{insig} = \sum_{i=C_b}^{N-1} |w(i)|^2$. According to the sorted sequence, it is clear that $\Delta = |w(C_b)|$. The average quantization error per significant coefficient is calculated based on the hypothesis that the *pdf* of the significant coefficients is flat within each quantization interval and thus the well-known approximation of a uniform distribution can be used, $\frac{E^{sig}}{C_b} = \frac{\Delta^2}{12}$. As shown in Figure 3.1, this explains why Mallat's model works better at *low* bitrates. This is because the histogram outside the central bin, Δ_L^o, is sufficiently flat, with coarse quantization, leading to an accurate approximation for E_{sig}. However, at *high* bitrates, the histogram outside Δ_H^o is not as flat, so that approximating by a uniform distribution within the interval may not be

accurate. Even though at high rates the quantization bins are small, they are not sufficiently small to make the uniform approximation sufficiently accurate. This can be verified in Figure 3.2 where the PSNR curve obtained from Mallat's model can be seen to deviate from the actual results at high bitrates.

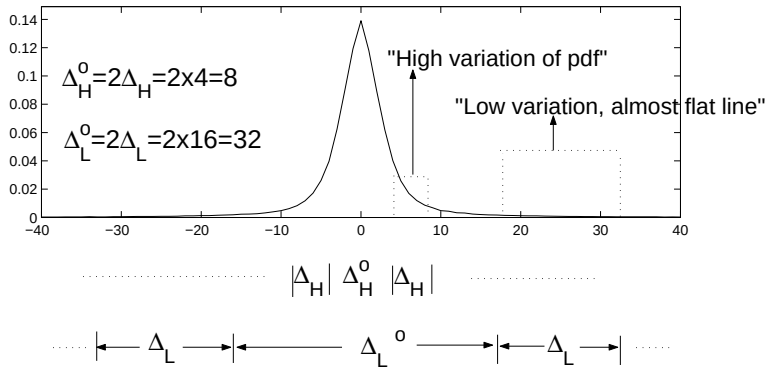


Figure 3.1: Normalized histogram of the wavelet coefficients of the gray-level Lena image. $\Delta_L = 16$ and $\Delta_H = 4$ are the sizes of final quantization bin and, $\Delta_L^o = 32$ and $\Delta_H^o = 8$ are the sizes of zero bin, at low and high bitrate (0.3 and 1.1 bps), respectively.

In our proposed model, we start by assuming that the *pdf* of the wavelet coefficients can be modeled as a Laplacian distribution [39]. To model E^{sig}, we start by computing the average error due to quantizing the significant coefficients in each quantization interval, i.e., coefficients located outside the zero bin, which have amplitude larger than the final threshold. Since SPIHT uses a uniform quantizer, all intervals are the same size and the reconstruction values are the midpoints of the interval. In this way, the error introduced in the i^{th} interval will be $\int_{i\Delta}^{(i+1)\Delta}(w-(i+\frac{1}{2})\Delta)^2 f_W(w)dw$ where $f_W(w) = \frac{\lambda}{2}e^{-\lambda|w|}$ represents a Laplacian *pdf* and λ is the Laplacian parameter. We average the error generated from all

intervals except the zero bin to obtain a new representation of the error for the significant coefficients, which is given by:

$$\begin{aligned} E^{sig} &= C_b\Big(2 \times \sum_{i=1}^{\frac{G}{2}} \int_{i\Delta}^{(i+1)\Delta} (w - (i+\frac{1}{2})\Delta)^2 f_W(w) dw\Big) \\ &= \frac{C_b e^{-K}(1 - e^{-\frac{KG}{2}})}{4\lambda^2(1-e^{-K})}\Big(K^2 - 4K + 8 - e^{-K}(K^2 + 4K + 8)\Big), \quad (3.2) \end{aligned}$$

where $K = \lambda\Delta$ and G is the number of quantization intervals besides the central bin. Note that we leave E^{insig} unchanged since there is no approximation in representing an error due to setting to zero the insignificant coefficients.

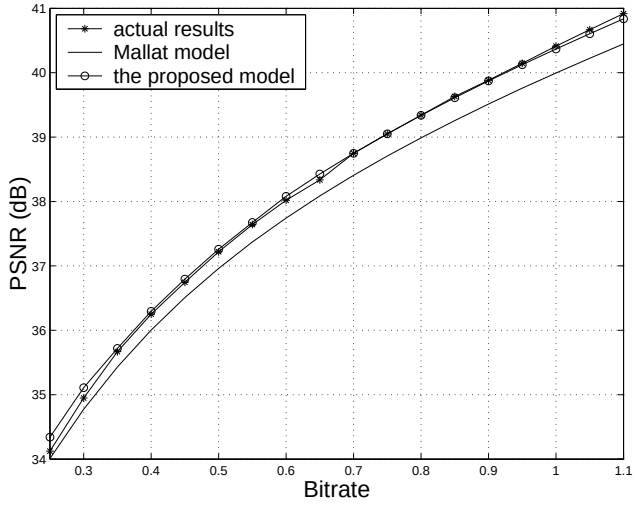


Figure 3.2: Comparison between the PSNRs of the gray-level Lena image obtained from actual experiment, those obtained from Mallat's model and those obtained from the proposed model [64]

With the new E^{sig}, the overall model is given by:

$$D(C_b) = \frac{\frac{C_b e^{-K}(1-e^{-\frac{KG}{2}})}{4\lambda^2(1-e^{-K})}\Big(K^2-4K+8-e^{-K}(K^2+4K+8)\Big)}{N} + \frac{\sum_{i=C_b}^{N-1}|X(i)|^2}{N}. \quad (3.3)$$

We validate the accuracy of the proposed model for different rate conditions in Figure 3.2. We can see that our model provides accurate results up to high bitrates. However, at very low bitrates (less than 0.3 bps), Mallat's model provides a better approximation. This comes from an inaccurate estimate of E^{sig} at low bitrates. This can be explained by the fact that the Laplacian model is initially designed to capture the characteristics of most coefficients, i.e., to best fit the empirical histogram of wavelet coefficients. Since most coefficients belong to the middle of histogram, the tail area of histogram (which has very low value of *pdf* and then is considered flat) might not be accurately estimated. However, Mallat's assumption fits well on this particular area since it assumes a flat histogram. As shown in Figure 3.2, although Mallat's model estimates better at very low bitrates, our proposed model is still able to give a reasonably accurate performance under these rate conditions as well. It is worth nothing that a hybrid model could be always used so that Mallat's model is used at low rates and the approximation of a uniform distribution distribution is used at high rates. However the decision could be based on determining how good of an approximation we have at a given

Δ, for a given image which may in turn be more complicated than the case where the only single model is used.

3.3 Analytical Model-based Bit Allocation for Optimization of Region of Interest Coding

Having established the proposed RD model, in this section we address an analytical model-based bit allocation for ROI coding. Here, we restrict ourselves to only two regions, for the sake of simplicity, but without loss of generality. First, we can write the distortion for each of the two regions: namely $D_1(C_1)$ and $D_2(C_2)$ representing region 1 and 2, respectively, by finding C_1 and C_2 and using them in Equation (3.3). Obviously, $D_{Total} = \frac{N_1}{N} D_1(C_1) + \frac{N_2}{N} D_2(C_2)$ where C_1 and C_2 are the number of significant coefficients and N_1 and N_2 are the total number of wavelet coefficients of region 1 and 2, respectively. It is clear that $C_b = \frac{NB}{6.6} = C_1 + C_2$ and $N = N_1 + N_2$. This can be simply generalized to obtain the distortion equations of the multiple regions. It is worth noting that given a bitrate budget B, the total number of significant coefficients, C_b, is a constant.

We now show how the proposed model can be applied to the divide-and-multiply method in ROI coding framework. We then illustrate the comparison of the overall RD operating points obtained by empirical technique, Mallat's model and the proposed model at a number of different psfs in Figure 3.3. To start the

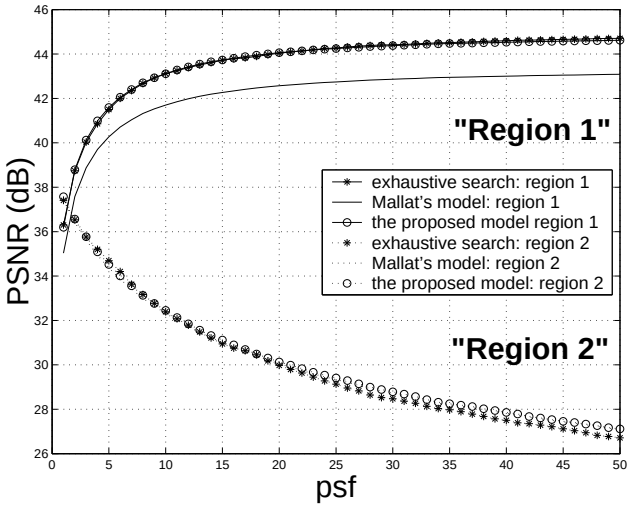


Figure 3.3: Comparison between the PSNRs obtained by exhaustive search, those obtained by Mallat's modeland those obtained by the proposed model. The gray-level Lena image, after dividing the wavelet coefficients outside the ROI by psf, is coded by a single SPIHT at rate 0.5 bps with a rectangular ROI of size 200×200 centered in the middle of the image.

analysis, the ROI is designated as region 1, and the rest of the image is called region 2. We divide the wavelet coefficients in region 2 by a given psf, so that the quantization bin of region 2 will be psf times larger than the quantization bin of region 1. It is worth noting that we use a normalized dividing factor so that the dividing factor for region 1 is always $psf = 1$. Clearly coefficients in region 2 are quantized with a final threshold $\hat{\Delta} = psf.\Delta$, while threshold Δ is used for region 1. For a given overall bitrate, as we increase the psf value, C_1 is increased and results in a decreasing distortion in region 1. But this increases the distortion in region 2 because C_2 must be reduced in order to keep $C_b = C_1 + C_2$.

As illustrated in Figure 3.3, we now explain why Mallat's model introduces an error in RD curves as compared to the benchmark results (the empirical curves), especially in the results of ROI (region 1), and why our proposed model can close this gap. Based on the divide-and-multiply method, as we increase the psf value, the coefficients in region 1 will be quantized with a finer quantizer than that in region 2. In other words, $\hat{\Delta}$ will get larger while Δ will get smaller, as the psf value increases, since $\hat{\Delta} = psf.\Delta$. Thus, the coefficients in region 1 are coded at high bitrate while those in region 2 are coded at low bitrate. Since the assumption of Mallat's model (the histogram outside the zero bin should be considered sufficiently flat) is not respected at high bitrate as illustrated in Figure 3.1, Mallat's model provides low accuracy in region 1. It is apparent that the results of the proposed model provide a more accurate RD curve, even when the coefficients of region 1 are assigned a very high bitrate, i.e., when a large psf value is used.

For region 2, both the proposed model and Mallat's model provide similar results as shown in Figure 3.3. Since coefficients of region 2 are coded at low bitrate, the histogram outside zero bin is sufficiently flat. It is worth noting that there is still a small gap between the empirical results and the proposed ones in *region 2*. Based on the experiments, which are done for a fixed bitrate B, as the psf changes, the *actual* number of significant coefficients, namely C_b^{act}, is *not* a

constant while the *analytical* number of significant coefficients, C_b, is constant[3] Instead C_b^{act} tends to decrease, when psf gets larger. Hence, as psf increases, C_b will be larger than the C_b^{act} and this will lead to a higher *analytical* PSNR than the *actual* PSNR. This affects mostly region 2 because the model expects to have the *additional* number of significant coefficients (i.e., $C_b - C_b^{act}$) assigned to the wavelet coefficients of region 2. Since this phenomenon occurs when very large psf is used, it results in very clear ROI but very blurred background. This is somewhat useless in the existing applications, and so the effect of this error is not significant.

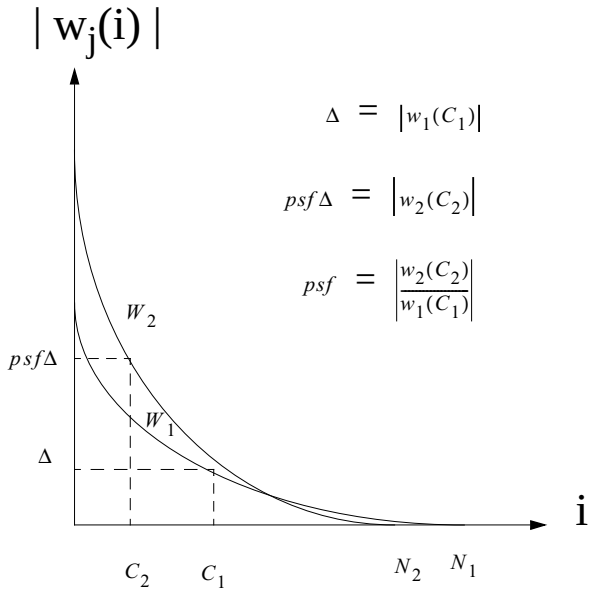


Figure 3.4: Curves of the sorted wavelet coefficients of each region, W_1 and W_2

[3]Recall that Mallat's model assumes that $C_b = \frac{NB}{6.6}$. This indicates that the number of significant coefficients depends only on the total number of coefficients N and the average bitrate B.

There are many possible distortion criteria that can be used to encode an image with ROI. Here, we consider the case where the goal is to determine the appropriate psf such that a desired ratio between distortion of non-ROI and ROI is satisfied. In other words, we seek to solve Equations (3.4)-(3.6) as shown below.

$$Ratio = \frac{D_2(C_2)}{D_1(C_1)} \tag{3.4}$$

$$C_b = C_1 + C_2 \tag{3.5}$$

$$psf = \left| \frac{w_2(C_2)}{w_1(C_1)} \right| \tag{3.6}$$

where w_1 and w_2 represents the sorted coefficients of ROI and non-ROI with quantization bins Δ and $psf \times \Delta$, respectively. $w_2(C_2)$ is the last coefficient sent for non-ROI and $w_1(C_1)$ is the last coefficient sent for ROI. This enables adjusting the relative importance of the regions as a $psf > 1$ indicates that only relatively larger coefficients are sent, and thus the bit rate required is lower. Equation (3.6) can be derived as shown in Figure 3.4. Using the proposed model of distortion as described in Section 3.2, we will be able to determine the psf without having to generate extensive RD data or without restricting the psf to take discrete values.

A simple calculation is performed over (3.4)-(3.6) to determine the appropriate C_1, C_2 and finally psf.

Algorithm: Analytical model-based bit allocation algorithm for ROI coding system

1. Start by computing the wavelet coefficients from the input image.
2. Split the wavelet coefficients into ROI and non-ROI, X_1 and X_2.
3. Sort the wavelet coefficients in each region in decreasing order to obtain the sorted sequence, W_1 and W_2.
4. Compute $D_1(C_1)$ and $D_2(C_2)$ as shown in Equation (3.3).
5. Determine the best C_1^* and C_2^* using Equations (3.4) and (3.5) for a given bitrate budget, B.
6. Determine the best psf_i^* from Equation (3.6) to obtain the optimal psf^*.
7. Divide the wavelet coefficients corresponding to the background by psf value computed from previous step.

3.4 Complexity Analysis

In terms of the complexity requirement, the proposed algorithm is much less complicated than the empirical-based method since it does not require that RD

characteristics be measured at each of the potential operating points. The most complicated process in the proposed algorithm is to sort the wavelet coefficients in monotonically descending order, e.g., using Quicksort algorithm, which requires $\mathcal{O}\ (N \log N)$ operations on average.

More specifically, given an image of size N pixels the model-based bit allocation algorithm requires: (i) $\mathcal{O}(N)$ operations for wavelet transform (ii) $\mathcal{O}(N)$ operations for polyphase transform, (iii) $\mathcal{O}(N \log(N))$ operations for Quicksort sorting process, (iv) $\mathcal{O}(N)$ operations for generating $D_i(C_i)$, (v) $\mathcal{O}(N)$ operations for determining the optimal $\{C_i^*\}$ (vi) $\mathcal{O}(1)$ operations for computing the optimal psf^* and (vii) $\mathcal{O}(N)$ operations for dividing psf^* to the wavelet coefficients in areas outside the ROI. Thus the total number of operations is $\mathcal{O}(N \log(N))$ operations. Based on this analysis, it is clear that the complexity of the proposed model-based analytical bit allocation approach is low (polynomial-time order).

3.5 Experimental Results and Discussion

In our experiments, we applied the distortion ratio criterion to three standard gray-level images of size 512×512 pixels (Lena, Boat and Lake). We validated our results by using different types of ROIs in different positions in each image: (i) a rectangular ROI of size 200×200 in the middle of the image, (ii) a cross-shaped ROI in the middle of the image, and (iii) an L-shaped ROI in the upper-left

corner of the image. We determine the psf obtained from our proposed model, denoted psf_{pro}, the one obtained using Mallat's model, denoted psf_{mal}, and the one obtained through an exhaustive search (i.e., selecting the best value among psf in the set $\{1, 1.1, 1.2, \ldots, 400\}$), which we denote psf_{emp}. Each of the psf values is chosen so as to target a desired distortion ratio between the ROI and the rest of the image, while minimizing the overall distortion, for a given total rate budget at 0.5 bps. We computed the variation between psf_{pro} and psf_{emp} and the variation between psf_{mal} and psf_{emp}. For each shape, we averaged the variations of the means (mean) and the standard deviations (std) over 3 images when the desired ratio varied from 1 to 10. We computed 3 statistical sets of data, which are: (i) psf values (ii) ROI distortions ($dist_{roi}$), and (iii) background distortions ($dist_{nroi}$). These are shown in Tables 3.1(a) and 3.1(b), based on our proposed model and on Mallat's model, respectively. Our results show clearly that our proposed model provides more accurately match the psf values and the distortions of the results obtained by optimization of empirical data as compared to those obtained by using Mallat's model.

To validate the subjective performance, we show the perceptual results between the reconstructed images using the psf_{emp} and the psf_{pro}. As illustrated in Figure 3.5, our approach provide an accurate estimate of the psf values to achieve the expected quality, and provides similar accuracy as the, significantly more complex, empirical-based method.

(a)

(b)

Figure 3.5: The reconstructed image with ROI of size 200×200 at the middle of the image when the empirical-based method (a) and the proposed algorithm (b) are applied to determine the psf value. The desired ratio is 50 and the total bitrate is fixed at 0.5 bps.

Shape	psf		$dist_{roi}$		$dist_{nroi}$	
	mean	std	mean	std	mean	std
rectangular	2.23%	1.70%	1.94%	1.64%	1.01%	0.91%
cross	2.83%	1.93%	1.95%	1.90%	1.70%	1.30%
L	2.30%	1.74%	1.74%	1.26%	1.30%	1.12%

(a)

Shape	psf		$dist_{roi}$		$dist_{nroi}$	
	mean	std	mean	std	mean	std
rectangular	6.75%	3.12%	2.99%	2.81%	3.45%	2.45%
cross	5.87%	3.01%	3.79%	2.47%	3.39%	2.16%
L	6.14%	3.50%	4.22%	3.75%	2.90%	1.81%

(b)

Table 3.1: Experimental results obtained by using (a) the proposed model and (b) Mallat's model

3.6 Conclusions

In this chapter we propose a novel analytical bit allocation based on using the proposed model to estimate RD characteristics in each region. We show that our analysis provides a bit allocation very close to that obtained by exhaustive search. We finally deploy our work in an ROI coding framework to demonstrate the efficiency of our model in a practical application.

Chapter 4

Channel Adaptive Multiple Description Coding for Image Transmission over Packet Loss Channels

In this chapter, we consider the problem of not only compressing data, as addressed in Chapter 2 and 3, but also robustly distributing the compressed data over unreliable channels. The main motivation is to provide an error-resilient coding algorithm by preserving some of the redundancy in the signal. This redundancy can be used to recover the data after channel losses. Additionally, the amount of redundancy can be easily adjusted as the characteristics of channel impairment change over time. [1]

[1] Work presented in this chapter was published in part in [63, 64, 62]

4.1 Introduction

One of the major challenges in achieving widespread delivery of real-time multimedia information comes in enabling such delivery over packet-based channels that are either subject to losses or suffer from variable delay. In this work we concentrate on error-prone transmission environments, such as the current best-effort Internet and the wireless IP networks, that are heavily subject to packet losses. These losses occur because packets are dropped as network queues fill up, or because network congestion delays their arrival to the receiver so that they can not be decoded before their scheduled playback time. In particular, in wireless networks, channel capacity may fluctuate due to several reasons such as multi-path fading, cochannel interference, changing distance between the base station and the mobile host, or network changes as a mobile terminal moves. These lead to time-varying transmission scenarios that result in a significant number of corrupted or lost packets. If there are no appropriate error and loss recovery mechanisms in place, significant quality degradation can be observed in the received multimedia signal.

One popular method dealing with lossy transmission environments has focused its attention on Layered Coding (LC) [43]. LC techniques have shown promising results in today's heterogeneous Internet because they enable quality adaptation, i.e., the base layer is sent first and then enhancement layers are

added to complete a full reconstruction [58], but when congestion occurs only the the base layer needs to be sent so that quality level can be adjusted to network conditions. The enhancement layer depends on the base layer and can not be decoded if the base layer is not received. The drawback of LC is that if base layer is lost during transmission the enhancement layers are useless. Therefore in LC the router or the intermediate nodes are expected to drop lower priority packets first (base layer packets are assigned higher priority than enhancement layer packets.) However, a common scenario in today's best-effort networks is that of transmission over a shared network where there are *no quality of service (QOS) guarantees* in the presence of packet losses. Thus, multimedia data is packetized for transmission but, given that there are no priorities, any packet could be lost during transmission under the no-priority-based infrastructure.

To improve the LC performance, several researchers have attempted to apply Automatic Repeat Request (ARQ) [26] techniques, which enable retransmission to be requested if needed for LC. Reliable protocols based on ARQ such as TCP are used in order to get the base layer across error free [75]. However, this may not be a practical solution if transmission is delay-constrained, especially when round-trip times (RTTs) are long. Retransmission of the lost packets can also contribute to network congestion, as it increases the network load.

Given that retransmission may not be desirable, an alternative is to use for reconstruction whatever is received at the destination. Traditional Forward Error

Correction (FEC) schemes have been used to recover from packet loss [40, 60]. These schemes operate by adding redundancy to the original data such that it can recover a certain number of losses at the destination. It however provides only a single level of protection. That is, once the packet loss rate is higher than the ability of FEC to recover the losses, the decoding quality will drop significantly (Cliff-effect). Thus no graceful degradation in reconstructed quality can be achieved if the number of losses exceeds the correction capacity of the code. Mohr *et al.* [48] has proposed to provide a graceful degradation scheme based on using FEC across the packets, Unequal Error Protection (UEP). All these FEC-based schemes would require a significant amount of interleaving and results in a delay especially when channel has bursty errors.

As an approach for source domain redundancy, Multiple Description Coding (MDC) has been proposed as an effective technique to provide robustness with graceful degradation under packet losses [19]. MDC can be seen as a forward error control technique in that the decoder reduces the error based only on the information that it received. MDC is particularly useful in scenarios where channels with unequal error protection are not available and retransmission is not desirable, as well as when there are strict delay constraints. Therefore, transmission of real-time media over the current Internet or wireless network infrastructure seems to be particularly well suited for MDC techniques since (i) differentiated

quality of service transmission has not been widely deployed, and (ii) RTTs can be significantly long.

In an MDC system the input signal is split into blocks and each block is represented by several descriptions. Then, these multiple descriptions of the source are sent to the receiver and it is assumed that random losses can affect each of the descriptions. Recovery of a particular block of data is possible as long as one of the packets carrying data of this block is received correctly. Unlike LC systems where layers are given different levels of importance, since enhancement layers are useless without the base layer, in an MDC system all descriptions are equally important, because each of the descriptions of the source can be decoded independently. When more than one description is received the overall quality can be improved with respect to having a single description.

The recent interest in MDC has led to proposal of several different MDC techniques [23, 24]. Examples include Multiple Description Scalar Quantizer (MDSQ) [70, 72], Multiple Description Transform Coding (MDTC) [82], and Unbalanced MDC (UMDC) [22, 74, 10, 2]. In an MDC system the basic trade-off is in the selection of the amount of redundancy. As can be expected, if high redundancy is chosen, the performance under severe error conditions will be good, while the performance in an error free environment will be significantly worse than that of a non-redundant coder at the same rate. Thus, the "right" level of redundancy may depend heavily on the specific target channel conditions. Given

that channel conditions in the Internet or wireless communication channels are time-varying, a useful target is to design MDC systems that can adapt to changing network conditions by adjusting their level of redundancy.

Several of the approaches mentioned above involve the design of specific transforms or quantizers that have to be matched to the desired level of protection [70, 72, 82, 24]. In these schemes, adapting to changing network conditions would entail having the encoder and the decoder both change the transform and/or quantizers. These approaches thus have limited ability to adapt to changing network/channel conditions. The MDC-FEC scheme [47, 55] is very popular due to its excellent coding performance. The idea of Unequal Error Protection (UEP) was applied to generate robust packet streams from a progressively compressed image [47, 55] such that we can determine which part of the multimedia stream require more protection. When network conditions change, re-generating the appropriate amount of FEC code is necessary to match the target channel conditions. However, both FEC computation and the interleaving are time-consuming processes which increases significantly the cost of achieving adaptation to channel/network conditions. One way to design these MDC approaches to account for varying communication conditions is to target the design for the worst case scenario. However this results in a performance penalty when the channel is error-free.

As an alternative, we concentrate on a class of MDC approaches first introduced by Jiang *et al.* [32] for image coding. These approaches are related to earlier work on audio coding [25]. In this technique, explicit redundancy is introduced, so that each input sample (for example each wavelet coefficient) is transmitted 2 times and coded with 2 different bitrates each time. This strategy has the drawback of leading to transmission of more samples than initially present in the source, and thus leads to inefficiency in cases of error-free transmission. However, this drawback is compensated by the extreme simplicity of the design, which relies on existing quantizers. As an example, wavelet-based MDC techniques [32, 45] can use the well-known SPIHT (Set Partitioning in Hierarchical Trees) coder [69] to encode the various descriptions in order to achieve excellent coding performance while providing robustness. This approach yields a simple design and our goal here is to enable efficient adaption of the levels of redundancy in order to increase both robustness capability and coding performance.

In this work we demonstrate how these explicit redundancy techniques have the additional advantage of providing very simple mechanisms to adapt to changing network conditions [63]. The key observation is that the level of redundancy can be selected by determining (i) the number of times a given sample (or wavelet coefficient) is transmitted, and (ii) how many bits should be used for each of the redundant representations. More specifically, we show how a *bit allocation* problem can be defined, where the goal is to choose the best distribution of redundancy

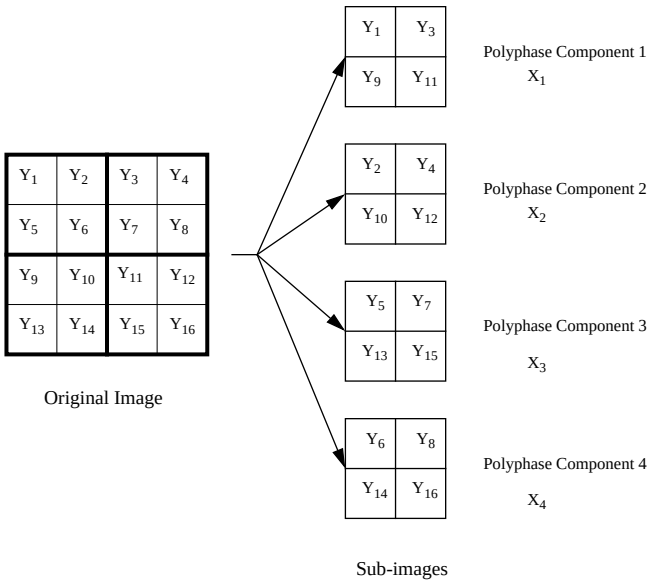


Figure 4.1: An example of polyphase transform when an original image, which is assumed to have a size of 4×4, is segmented into 4 blocks of size 2×2 where Y_i represents the i^{th} pixel, $i = 1, \ldots, 16$. A polyphase component is obtained by picking from each subblock a pixel appearing in the same relative position.

for a given packet loss rate. We assume that the packet loss rate can be monitored and estimated via the feedback channel, e.g., RTCP packet. We provide techniques to solve this problem and show how indeed different loss rates require different levels of redundancy. Note that by using bit allocation to determine the level of redundancy, not only the encoder can adjust itself in a simple manner, but in addition the decoder can handle packets with different levels of redundancy without requiring any changes to its structure (e.g. the same transform, entropy coding, etc can be used).

The proposed MDC technique generates the various descriptions through a polyphase transform. To obtain polyphase components for the case of a scalar

source, this polyphase-based MDC divide this source, say a 1-dimensional sequence $\{Z_1, \ldots, Z_n\}$ (for simplicity assuming that n is an even number), into even and odd samples, i.e., $\{Z_2, Z_4, \ldots, Z_n\}$ and $\{Z_1, Z_3, \ldots, Z_{n-1}\}$ (or more sets if more than two descriptions are required), and will compress each sample using two different quantization scales (coarse and fine). For an $n \times n$ image or 2-dimensional signal source, $\{Y_{(1,1)}, Y_{(1,2)}, \ldots, Y_{(n,n)}\}$, and for 4 polyphase components, $Y_{(i,j)}$ belongs to polyphase component $i \bmod 2 + 2 \times (j \bmod 2)$. An example of the polyphase transform for an original image of size 4×4 composed of 16 pixels $\{Y_1, \ldots, Y_{16}\}$ as a 2-dimensional source is provided to illustrate the definition in Figure 4.1. Then groups of samples are transmitted where a set of coarsely quantized odd samples is combined with a set of finely quantized even samples (and vice versa, i.e. fine odd with coarse even.)

The MDC system block diagram at the transmitter is provided in Figure 4.2 in the general case when S descriptions (DC_i for $i = 0, \ldots, S-1$) are used, based on S polyphase components (X_j for $j = 0, \ldots, S-1$). In this case, each description contains M polyphase components from fine level to coarse level, each corresponding to different input samples, i.e., M polyphase components are encoded with different quantization steps from fine to coarse. We call each packet as described here a *description*, while we call each redundant version of a polyphase component a *copy*. Each copy is coded with a different quantizer. For example, in Figure 4.2 we show how M copies of a given polyphase component

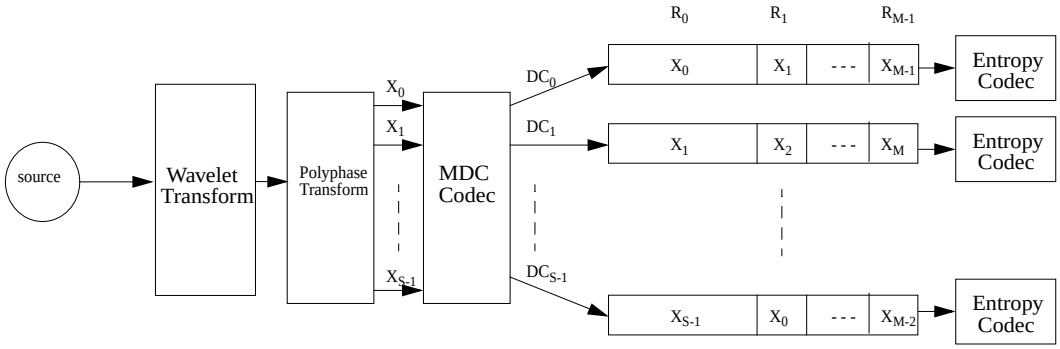


Figure 4.2: MDC system block diagram: S descriptions are generated by obtaining S polyphase components from the original signal. For each polyphase component M copies are transmitted. Each description carries the primary copy of one polyphase component, e.g., X_0 in DC_0, as well as redundant copies of some of the other polyphase components.

are sent, each in one description, with rates ranging from R_0 to R_{M-1}. A given description contains the primary copy of one polyphase component, i.e., the copy coded at highest quality, along with lower resolution copies of other polyphase components. Note that our definition of description is consistent with that of traditional MDC approaches, i.e., the more descriptions are received the better the quality of the decoded signal.

The decoder operates by gathering the available information for each polyphase component and then selecting for each polyphase component its highest quality copy to be used in the decoding; the remaining copies are discarded. For example, referring again to Figure 4.2, if only DC_1 is lost then all polyphase components except X_1 will be decoded at their highest quality, R_0, while X_1 will be decoded with its highest received quality, R_1.

Obviously, as the packet losses increase, the chances that the highest quality copy of a polyphase component reaches the receiver diminish. In that situation, it would be better to distribute bits more evenly among all the copies. Therefore, the bit allocation we address in this work is as follows. Given a known packet loss rate, our goal is to determine (i) the optimal number of copies of a polyphase component and (ii) the amount of bits to assign to each copy. In other words, our goal is to determine the number of polyphase components to be included in a description, M, and the number of bits to be used for each polyphase component in a description, R_0, R_1, ..., R_{M-1}, such that the overall expected distortion at the receiver due to both coding and description loss is minimized for a given target packet loss rate, and for a given fixed rate budget. It has to be pointed out that, unlike other recent MDC approaches [32, 70, 82, 24], this polyphase-based MDC system can easily incorporate more than 2 descriptions. This is useful given that use of only 2 descriptions may be too limited for multicast applications, for which more than two levels of reconstruction may be required.

It is worth noting that the data generation needed to obtain the Rate-Distortion (RD) data to be used in the MDC bit allocation is a time-consuming process. Although the MDC system has a simple design and provides fast adaptation in redundancy during a time-varying transmission, the complexity can be dominated by the data generation process. We therefore introduce a novel RD model to simplify the data generation needed for MDC bit allocation. In this chapter,

we use our efficient RD model proposed in Chapter 3. Recall that the proposed RD model is an extended version of the RD model originally introduced by Mallat and Falzon [42]. While the original RD model has been proposed to operate only at low bitrates, we extend it so that it can operate at any bitrate. With the proposed efficient RD model we enable an MDC approach that is capable of adapting to changing network conditions.

We finally extend our MDC system so that it is able to provide different levels of protection to different parts of an image. That is, we add more redundancy to the Region of Interest (ROI) than to the rest of the image. A similar approach of combining ROI coding and MDC approach has been introduced by Miguel *et al.* [44]. However they use a different bit allocation technique without any optimization.

This chapter is organized as follows. In Section 4.2 we formalize the problem. An algorithm using Lagrangian optimization is introduced in Section 4.3 where each polyphase component included in each description is separately coded. To exploit the correlation lost by the independent coding, in Section 4.4 we enhance the coding performance by coding all polyphase components included in one description together. Under this dependent coding environment, we use our proposed RD model to eliminate the complexity due to data generation process and use it to perform the bit allocation for our MDC system. We address the problem of protecting a region of interest by using the proposed MDC system in

Section 4.5. In Section 4.6, experimental results are provided as demonstration of the validity of our analyses. Finally, conclusion of this work is discussed in Section 4.7.

4.2 Problem Formulation

As outlined in the introduction, our goal is to determine what the right amount of redundancy is for a given packet loss. Given that S polyphase components are generated and S packets are transmitted for each basic coding unit, our goal is to determine M, the number of copies of each polyphase components to be transmitted. Obviously, if $M = 1$ we introduce no redundancy, and thus any packet loss will result in a polyphase component being lost. Conversely if $M = S$ the maximum redundancy is introduced, and therefore we will be able to reconstruct all the polyphase components of the input, albeit with different levels of quality, as long as at least one packet out of S is received. A related issue is that of determining, once M has been chosen, the number of bits to be assigned to each copy, i.e., R_0, R_1, ..., R_{M-1}. We will show how these two decisions can be made jointly in the process of performing a bit allocation, by reducing M to $M - k$ if R_{M-k}, ..., R_{M-1} are assigned a negative (or zero) number of bits in the bit allocation.

Clearly, the "right" choice for M will depend on the channel conditions, so that high redundancy is to be expected when packet losses are high and low redundancy should be used under low packet loss conditions. Since increasing the number of polyphase components in a description lowers the coding efficiency, while increasing the robustness to errors, it is to be expected that at each loss rate an optimal redundancy can be determined. The goal in our bit allocation problem is to minimize the *expected distortion* at the receiver, after taking into account the effect of packet losses.

We start by evaluating the expected distortion at the receiver as a function of the packet loss rate. Assume a total bit rate budget B has been fixed, and assume a known probability that a description is considered lost, P, which is independent of the level of redundancy in each description (we assume that the packet size is fixed in all the scenarios we compare.) Given our total number of polyphase components, S (which is equal to the number of descriptions) and given the total number of samples N, each polyphase component provides information for $\frac{N}{S}$ samples. Note that of the M polyphase component copies sent in one description one of them will be primary copy and the remaining $M-1$ copies will be redundant of polyphase components whose primary data is transmitted in another description. Given a bitrate budget of B bits per sample on average, we have that the description size is $\frac{BN}{S}$ bits, which does not depend on the number of polyphase components in a description.

The primary copy is coded with rate R_0 while the other copies are coded with rates $R_1, \ldots, R_{M-1}$ with $R_m \geq R_n$ for $m < n$ and $m, n = 0, \ldots, M-1$. It is worth noting that our analysis in this work is description-based, i.e., we assume that each description is contained in a single packet. However similar analyses can be extended to a more practical packet-based framework, where the packet size is smaller than the size of a description. For instance, if we need to operate with small packet sizes, instead of considering the whole original image as the input in our analysis, we can divide the image into smaller blocks, and then extract the polyphase components in each of these blocks.

In order to estimate the distortion at the receiver we need to specify the decoding algorithm. In general, the decoder will receive more than one copy of each polyphase component. Then, the decoder will select among all those copies the one with the highest resolution (i.e., higher rate R_i) and will use it for reconstruction, while discarding all other copies. In the worst case, when no copy of a polyphase component is received, the decoder will use the mean value of this polyphase component. This is sent as side information which would not require much overhead.

The expected distortion, $E[D]$, which we seek to minimize, statistically measures the reconstructed quality at the receiver. Given our proposed MDC approach, the distortion incurred for a given sample will be that corresponding to the highest quality copy received for that sample. In the worst case, if all

copies are lost, the distortion will be the variance of the corresponding polyphase component.

The expected distortion can be derived by determining the probabilities that the best copy for a given polyphase component is the one with index i. Note that we assume that the packet structures are identical and therefore it follows that these probabilities will be the same for each polyphase component. Denote P_i for $i = 0, \ldots, M$, as the probability that copy i is the highest quality one received for the polyphase component under consideration. P_M is the probability that none of the copies is received (since there are only M copies, with indices 0 to $M-1$.) P_0 is the probability that the primary copy (with high rate equal to R_0) is received. Denote $D_{ji}(R_i)$ as the distortion associated with copy i of polyphase component j, coded with R_i bps. Then the expected distortion for polyphase component j will be $\sum_{i=0}^{M-1} P_i D_{ji}(R_i)$.

Let us compute the probabilities. Consider the best scenario where, for instance, the description containing polyphase component j as a primary data (i.e., it is coded with the highest bit rate R_0) is received. Given a packet loss probability P, the probability of this case is equal to the sum of the probabilities of all scenarios where copy 0 of polyphase component j is received: $P_0 = 1 - P$. That is, we must ensure that the packet carrying that copy arrives, independently of whether the others are received.

In general, the probability of receiving copy i as the one with the highest quality for polyphase component j will be the probability of receiving the corresponding packet correctly, but losing i higher quality copies, i.e. those coded at rates R_0 through R_{i-1}. This probability can be expressed as: $P_i = (1-P)P^i$.

Finally, none of the copies of polyphase component j are received if all the corresponding M packets are lost. Therefore the corresponding probability can be written as: $P_M = P^M$.

The total expected distortion can be computed by adding the distortions of each polyphase component divided by total number of polyphase components, S. Thus, the expected distortion is

$$E[D] = \frac{1}{S}\Big(\Big[\sum_{j=0}^{S-1}\sigma_j^2\Big]P^M + (1-P)\Big[\sum_{i=0}^{M-1}\Big\{P^i\sum_{j=0}^{S-1}D_{ji}(R_i)\Big\}\Big]\Big), \qquad (4.1)$$

where $D_{ji}(R_i)$ is the distortion of polyphase component j when R_i bits are used, and σ_j^2 is its distortion when all M copies are lost.

4.3 Optimization

When the expected distortion can be computed as above, our goal then is to find the best bit allocation R_i for each of the descriptions, which also leads to finding the best M for a given P. Finding the best bit allocation can be stated as a constrained optimization problem, where the R_i is selected to minimize

the distortion from Equation (4.1) subject to a total budget constraint, i.e., the average bit rate per sample has to be equal to the budget, B: $\sum_{i=0}^{M-1} R_i = B$.

Lagrangian optimization techniques [20, 73] can be used to solve this problem by introducing a cost function, J, where a Lagrange multiplier, $\lambda \geq 0$, is used to trade-off rate and distortion. This leads to an unconstrained minimization of the cost function,

$$J = E[D] + \lambda\Big(\sum_{i=0}^{M-1} R_i - B\Big). \tag{4.2}$$

The optimization process can be summarized as follows. First, we set the number of polyphase components in each description, M, to be equal to the maximum possible, i.e., S, which provides the maximum level of protection. Note that while each polyphase component may in general have different RD characteristics, here we are assuming that all the packets are structured in the same way so that copy i of any polyphase component will always be allocated R_i bits. In this work, we further assume that the RD characteristics are the same for all polyphase components. This bit allocation can be found based on either closed form RD models or empirical RD data. In both cases we use the Lagrangian optimization technique.

Consider first the optimization based on a closed form model. In other words we assume that the RD characteristics of the source data is available. As an

example, we consider the case of an i.i.d. zero mean Gaussian random source, with RD characteristic [11] given by: $D(R) = \sigma^2 2^{-2R}$.

Given that performing a polyphase transform with uniform sub-sampling will not change the rate-distortion characteristics of such a memoryless Gaussian random source, we can assume that each polyphase component has the same rate distortion characteristic (i.e. from Equation (4.1), $\sigma_j^2 = \sigma^2$ and $D_{ji}(R_i) = D_i(R_i)$ for all $j = 0, \ldots, S-1$). Thus, the expected distortion per sample will be

$$E[D] = \sigma^2 P^M + \sigma^2 (1-P) \sum_{i=0}^{M-1} P^i 2^{-2R_i}. \tag{4.3}$$

Hence, our objective can be stated as follows:

$$\begin{aligned} Minimize \quad E[D] &= \sigma^2 P^M + \sigma^2 (1-P) \sum_{i=0}^{M-1} P^i 2^{-2R_i} \\ subject\ to \quad B &= \sum_{i=0}^{M-1} R_i. \end{aligned} \tag{4.4}$$

To minimize the average distortion in the presence of channel failures, we introduce an unconstrained cost function as shown below and then differentiate this constrained cost function with respect to R_i to determine λ^* and R_i^*. Finally optimal bit allocation for each polyphase component in a description is derived and expressed as shown below.

$$J = \sigma^2 P^M + \sigma^2 (1-P) \sum_{i=0}^{M-1} P^i 2^{-2R_i} + \lambda \Big(\sum_{i=0}^{M-1} R_i - B \Big) \tag{4.5}$$

$$R_i^* = \frac{B}{M} + \frac{1}{2}\Big(i - \frac{M-1}{2}\Big)\log_2\Big(P\Big) \tag{4.6}$$

It is worth noting that the optimal bit allocation can be computed mathematically and directly because a closed form of a rate-distortion function of each polyphase component is explicitly assumed.

The solutions obtained above may result in some of the R_i^* to be negative. If this is the case and, say, R_k^* through R_{M-1}^* are all negative, we would restart the optimization choosing $M = k$. This is illustrated by the flow diagram in Figure 4.3.

The same Lagrangian algorithm can also be applied based on empirical data of the source. Here we first split the source (e.g., an image in the experiments section) into S polyphase components using the polyphase transform. Then we independently code each polyphase component, and measure the RD values that can be achieved. Given a packet loss rate P and budget B, we iterate over a non-negative λ until we find one such that the budget is met. One particular operating point is that where the rate is zero and the distortion is the variance of the source. Thus, in the resulting bit allocation (R_i^* for $i = 0, \ldots, S-1$), if some of the copies have been allocated zero bits that means that those copies should not be transmitted and therefore that the M should be smaller. For example, if the optimal bit allocation results indicate that the rates $R_k, \ldots R_{S-1}$ should

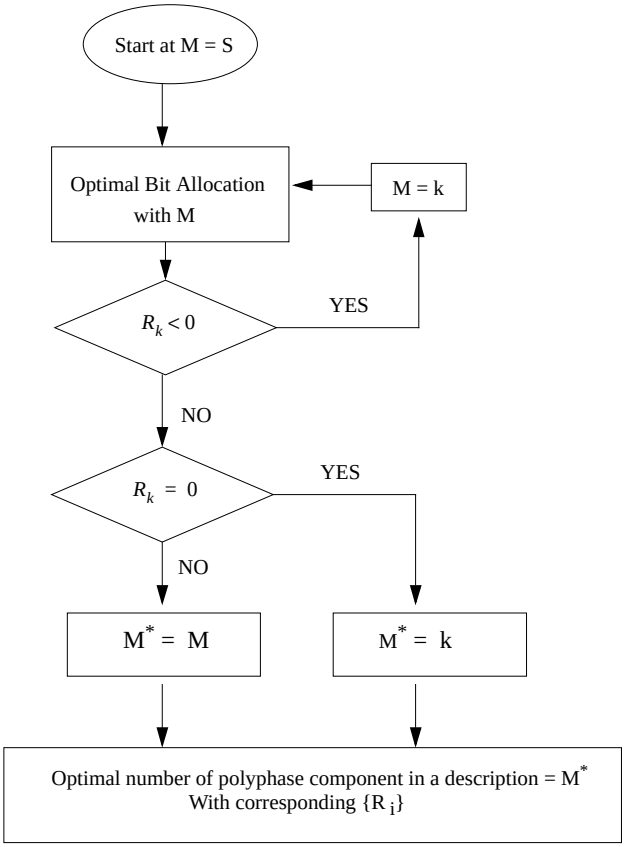


Figure 4.3: The algorithm flow diagram

be zero, then that means that the optimal level of redundancy is $M = k$. Note that this process is different from the previous optimization method since no close form model (e.g., the i.i.d. zero mean Gaussian random source) is assumed and thus we need to search for the best Lagrange multiplier and the optimal bit allocation. Moreover, the solutions obtained above can not be such that some of the R_i^* are negative.

It has to be pointed out that in the empirical-base optimization scheme (when the closed form model of input source is *not* available), the complexity from the RD data generation process significantly dominates the complexity from the optimization process, i.e., finding all $D_{ji}(R_i)$ points for the problem is much more complex than finding the optimal solution, given all the possible R-D operating points. In particular, in next section, instead of coding each polyphase component separately, we propose to code all polyphase components inside the same description together such that we can effectively exploit the correlation between them. In that way, the distortion for each polyphase component does not only depend on its bitrate, but also depends on the choices of the bitrates of other polyphase components inside the same description. In this case, finding all $D_{ji}(R_0, R_1, \ldots, R_{M-1})$ points is needed for the RD data population. Clearly this dependent coding method results in much higher complexity as compared to the previous independent coding method. With this computational burden, a channel-adaptive MDC would be difficult to achieve in real time. Therefore, we

are interested in model-based bit allocation methods that will avoid the need to generate real RD data while preserving the optimality. However it is clear that i.i.d Gaussian assumption can not be assumed for every input image. Therefore it would be useful to have an RD model that can work on arbitrary input images. In the next section, we introduce a bit allocation algorithm based on our proposed RD model in Chapter 3. We show how our method can be used to solve this dependent bit allocation problem for the proposed MDC system and show that the complexity is significantly reduced.

4.4 Analytical Model-based Bit Allocation for MDC

In the previous section, each polyphase component included in each description is encoded separately. Clearly, the system does not fully exploit the correlation among the polyphase components inside the same description. It would be preferable to code *all* polyphase components included in a description together such that the correlation can be effectively exploited. However this introduces the problem of how to effectively use the desired number of bits, R_i, on the different polyphase components included in a description, such that given the packet loss rate, P, the expected distortion will be minimized.

In this section, we improve our MDC system by using a divide-and-multiply method used in Region of Interest (ROI) Coding systems that was introduced in the previous chapter. Recall that in ROI coding different numbers of bits are allocated to different regions in an image coded with a progressive wavelet coder such as SPIHT [69] or JPEG2000 [6]. The wavelet coefficients are divided by different factors before coding to enable different bit allocations to different regions, because the coefficients in each region are refined at different speeds.

Based on the ROI coding idea, the divide-and-multiply method can be used to introduce redundancy in each description [64]. That is, we divide each redundant version of a polyphase component by a psf (Priority Scaling Factor). In this way different numbers of bits will be efficiently distributed among the polyphase components in each description as shown in Figure 4.4. At the encoder we divide the polyphase components with different psf values, $psf_0, psf_1, \ldots, psf_{S-1}$. More specifically, corresponding to the original MDC structure in Figure 4.2, the primary copy is divided by psf_0 while the other copies are divided by psf_1, $\ldots$, psf_{S-1} with $psf_m \leq psf_n$ for $m < n$ and $m, n = 0, \ldots, S-1$. At the decoder, we multiply the psf values used at the encoder to the corresponding polyphase component. In this way, different numbers of bits will be assigned to different polyphase components during the coding process via the bitplane-by-bitplane successive refinement.

Similar to the empirical-based optimization mentioned in Section 4.3 where if $M < S$, $R_k^* = 0$ for $k = M, \ldots, S-1$, a large value of psf_k^* for $k = M, \ldots, S-1$ is expected to be used in the MDC system. In this way, all bits will be distributed among the primary copy and the next $M-1$ copies and no bits will be used by the remaining copies.

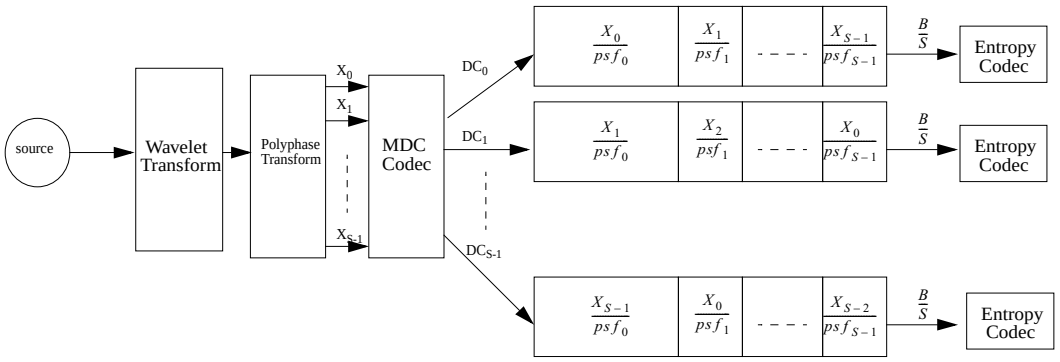


Figure 4.4: Block diagram of S-description system using psf

In general the polyphase transform is used to generate regions such that each region will have the same size, $N_j = \frac{N}{S}$ for $j = 0, \ldots, S-1$, and similar properties. The analytical psf technique has the advantage of being simple (a division operation) and flexible (the levels of redundancy can be easily changed).

To preserve our total bit rate budget we code each description at bitrate $\frac{B}{S}$. It is worth noting that we use a normalized dividing factor. Without loss of generality assume that $psf_0 = 1$. This results in no loss of generality because the rate is controlled through the bitplane coding. Define $\mathbf{psf} = psf_0, psf_1, \ldots, psf_{S-1}$. Denote $D_{ji}(\mathbf{psf})$ as the distortion associated with copy i of polyphase component

j for $i = 0, \ldots, S-1$ and $j = 0, \ldots, S-1$. Under this dependent coding environment, our notation, $D_{ji}(\mathbf{psf})$ for the distortion of the i^{th} copy of polyphase component j indicates that psf choices for other copies in the same description affect the result. As derived in Section 4.2, the expected distortion can be computed as,

$$D_{avg}(\mathbf{psf}) = \frac{1}{S}\Big[\sum_{j=0}^{S-1}\sigma_j^2\Big]P^S + \frac{1-P}{S}\Big[\sum_{i=0}^{S-1}P^i\sum_{j=0}^{S-1}D_{ji}(\mathbf{psf})\Big]. \tag{4.7}$$

Now our goal is to find the best $\mathbf{psf}^* = \{psf_0^*, psf_1^*, \ldots, psf_{S-1}^*\}$ given the probability of packet loss P. Therefore, if we assume that we have access to the distortion curve $D_{ji}(\mathbf{psf})$, searching all over the RD operating points should yield the best **psf**. However, data generation is normally a time-consuming process [57]. That is, the encoder has to generate all RD operating points before starting the optimization process, i.e., the rate and distortion characteristics have to be measured at each of the potential operating points. For example, a design based on empirical data could start by measuring overall image RD data at a number of different psf values, and then proceed to select the optimum or the most appropriate psf value for a given criteria. There are two major drawbacks of this empirical-based method. First, it is obvious that the technique is limited in that the solution will have to be one of the discrete quantizer sets, so that

optimality may suffer if a bad choice was made of those discrete quantizers. Second, collecting the RD data may be complex. The more admissible quantizers are used, the more time will be needed in the RD data generation process. Therefore, it would be useful to design a model-based bit allocation algorithm that can both efficiently work on arbitrary input images and enable fine granularity in the selection of arbitrary admissible quantizer sets. Next, we introduce a novel RD model for the purpose of RD data generation and show how to use it to determine the best **psf** to use for the MDC bit allocation given a packet loss rate and a bitrate budget.

In the MDC system with S descriptions, the average distortion can be simplified from Equation (4.7), by using the distortion model in Equation (3.3), with the new parameters, C_i, the number of significant coefficients for copy i of the polyphase component. It is worth noting that C_i is a function of **psf**, i.e., C_i depends on the coding choice. To compute the distortion for copy i of polyphase component j, $D_{ji}(C_i)$, for $i = 0, \ldots, N-1$ and $j = 0, \ldots, N-1$, the source is first decomposed into S polyphase components. Then we independently sort the wavelet coefficients in each polyphase component to obtain S sorted sequences, $W_j = \{w_j(0), \ldots, w_j(\frac{N}{S} - 1)\}$ for $j = 0, \ldots, S-1$.

$$D_{avg}(C_0, C_1, \ldots, C_{S-1}) = \frac{1}{S}\Big[\sum_{j=0}^{S-1}\sigma_j^2\Big]P^S + \frac{1-P}{S}\Big[\sum_{i=0}^{S-1}P^i\sum_{j=0}^{S-1}D_{ji}(C_i)\Big]$$

$$C_B = \sum_{i=0}^{S-1} C_i \tag{4.8}$$

where $D_{ji}(C_i)$ represents the distortion of polyphase component j that is divided by the psf_i.

Therefore, given the probability of packet loss, P, we can determine the optimal C_i^* by using Lagrangian optimization techniques [73]. The optimal C_i^* are determined from the optimization process and then each psf_i^* can be determined as

$$psf_i^* = \frac{|w_i(C_i^*)|}{|w_0(C_0^*)|} \tag{4.9}$$

where $w_i(c_i^*)$ is the last coefficient sent for copy i and $w_0(c_0^*)$ is the last coefficient sent for copy 0. This enables adjusting the relative importance of the copies as a $psf > 1$ indicates that only relatively larger coefficients are sent, and thus bitrate required is lower.

Without loss of generality, we give an example for a 2-description system as shown in Figure 4.5. W_0 and W_1 represent the sorted coefficients of polyphase component 0 and 1 with the different quantization bins at Δ and $psf \times \Delta$, respectively. The psf value can be determined as shown in Figure 4.5. For an S-description system, it is therefore clear that the psf_i^* can be determined as shown in Equation (4.9) and used as the input parameter to the MDC codec afterwards.

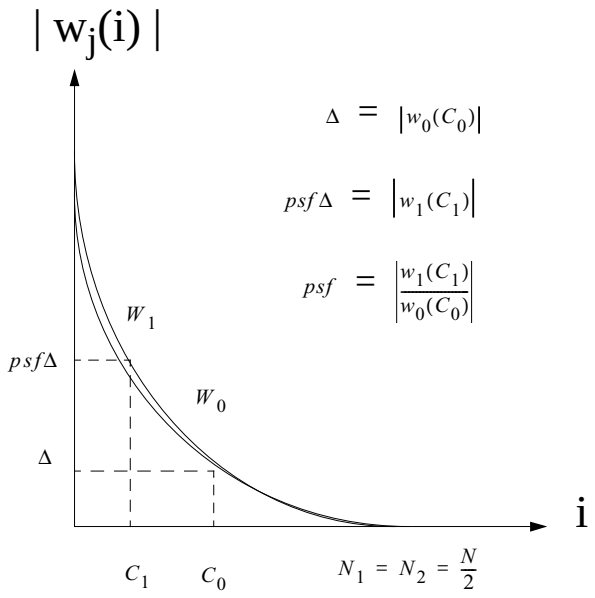


Figure 4.5: Curves of the sorted wavelet coefficients of each polyphase component, W_0 and W_1

Algorithm: Analytical model-based bit allocation algorithm for MDC system

1. Start by computing wavelet coefficients from an input image.

2. Split the wavelet coefficients into S polyphase components, X_j for $j = 0, \ldots, S-1$: $X_j = \{x_j(0), \ldots, x_j(\frac{N}{S} - 1)\}$.

3. Sort the wavelet coefficients in each polyphase component in decreasing order to obtain the sorted sequences, W_j for $j = 0, \ldots, S-1$: $W_j = \{w_j(0), \ldots, w_j(\frac{N}{S} - 1)\}$.

4. Compute the $D_{ji}(C_i)$ in the same way as shown in Equation (3.3) and use it in Equation (4.8).

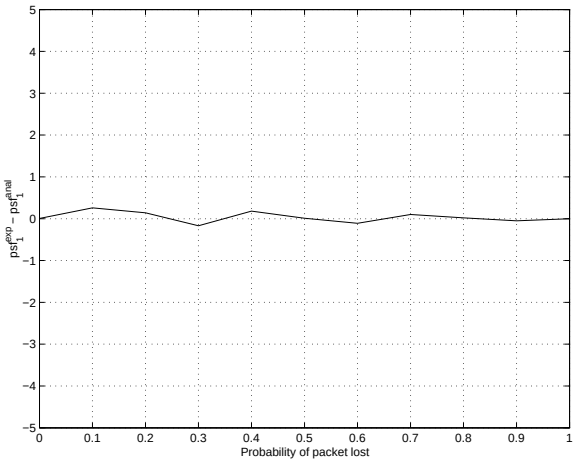
5. Perform Lagrangian optimization by minimizing Equation (4.8) with the constraint $C_B = \sum_{i=0}^{S-1} C_i$ to determine the best C_i^* for $i = 0, \ldots, S-1$, for a given packet loss rate, P, and a given bitrate budget, B.

6. Determine the best psf_i^* from Equation (4.9) to obtain the optimal $\mathbf{psf}^*$.

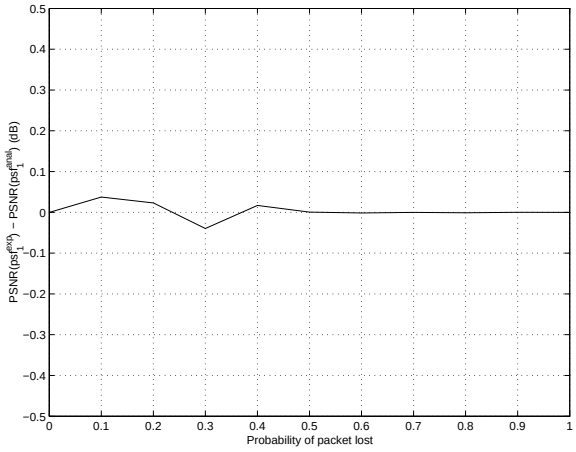
7. Divide the wavelet coefficients for each polyphase components in a description as shown in Figure 4.4 by the appropriate psf values computed from Step 6.

8. If there is a change in channel conditions, go back to Step 5 to re-compute the new psf value for a new packet loss rate and a bitrate budget.

To validate the accuracy of the proposed algorithm, a comparison between the empirical-based and model-based psf values for a 2-description system is presented in Figures 4.6(a) and (b) where various different channel conditions are assumed. It is worth noting that this is a comparison when empirical or model RD curves are used with different packet loss. The empirical-based psf_1, denoted by psf_1^{emp}, is obtained from full search of all admissible psf_1 of $\{1, 1.1, 1.2, \ldots, 400\}$ while the model-based psf_1, denoted by psf_1^{anl}, is determined from our analysis as shown in Figure 4.5. Note that psf_0 is equal to 1. From Figure 4.6 it is clear that the proposed scheme provides accurate estimates of the psf values.



(a)



(b)

Figure 4.6: (a) Difference between the empirical psf_1 and analytical psf_1 and (b) difference between the PSNR results using the empirical psf_1 and the analytical psf_1 for gray-level Lena image of size 512 $\times$ 512. MDC system generates 2 descriptions with a total bitrate at 0.5 bps.

4.5 Local vs Global Protection using ROI Coding with MDC System against Packet Loss

The proposed MDC approach can be extended to ROI coding to provide higher protection to more important portions of the image. In this way, higher quality can be achieved for the ROI than the rest of the image, i.e., non-ROI (e.g., background). We first divide the non-ROI coefficients for each polyphase component in a description with different dividing factors $r_0, r_1, \ldots, r_{S-1}$. This is equivalent to dividing the ROI and non-ROI coefficients for each polyphase component. For example, in the first description, DC_0, we use $psf_0, psf_1, \ldots, psf_{S-1}$ for the ROI and $psf_0r_0, psf_1r_1, \ldots, psf_{S-1}r_{S-1}$ for the non-ROI as shown in Figure 4.7. At the decoder we multiply by the corresponding dividing factor the reconstructed coefficients before the inverse transform is performed. In this way more bits will be distributed and better quality will be achieved at ROI than at non-ROI. Thus MDC can be viewed as a way to **globally** protect the more important data in the entire image while ROI coding is a method to **locally** protect more strongly information in some parts of the image. A similar approach has been introduced by Miguel *et al.* [44] to solve this problem. However that work uses a different bit allocation technique and no criteria are presented to select the dividing factor, while in this work the dividing factor for ROI coding can be selected based on both the distortion criteria and the minimization of the average distortion.

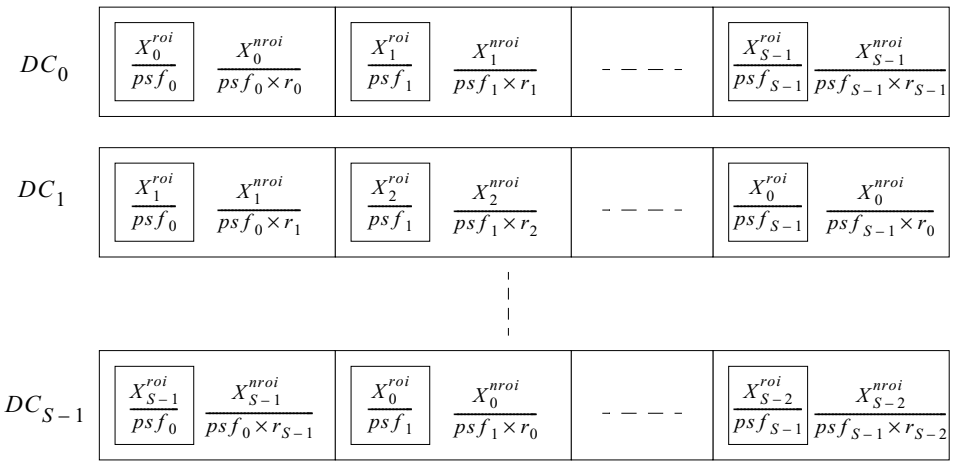


Figure 4.7: Block diagram of S-description system with the ROI coding. The ROI is represented as a small rectangular box inside each polyphase component.

There are many possible distortion criteria that can be used to encode an image containing an ROI to provide different levels of protection to each portion of the image. Here, we consider the case where the goal is to determine the appropriate psf values to be assigned to each polyphase component, i.e., $r_0, r_1, \ldots, r_{S-1}$, such that a desired ratio between distortion of non-ROI and ROI is satisfied where r_j is the psf value for the divide-and-multiply method for ROI coding in each copy of polyphase component j in a description. We can write the distortion for ROI and non-ROI of polyphase component j as $D_{ji}^{roi}(C_i^{roi})$ and $D_{ji}^{nroi}(C_i^{nroi})$, respectively.

Again, assuming that the characteristics of each polyphase component will be approximately the same, the distortion function for each description will be the same, i.e., $D_{ji}^{roi}(C_i^{roi}) \approx D_i^{roi}(C_i^{roi})$ and $D_{ji}^{nroi}(C_i^{nroi}) \approx D_i^{nroi}(C_i^{nroi})$. Given

that C_i^* is computed as described in the previous section, a simple calculation can be performed to determine the appropriate C_i^{roi} and C_i^{nroi} and finally r_i^* is computed, as shown below. Note that $w_i^{nroi}(C_i^{nroi})$ and $w_i^{roi}(C_i^{roi})$ are the sorted non-ROI and ROI coefficients of the polyphase component i, respectively.

$$Ratio = \frac{D_i^{roi}(C_i^{roi})}{D_i^{nroi}(C_i^{nroi})} \tag{4.10}$$

$$C_i^* = C_i^{roi} + C_i^{nroi} \tag{4.11}$$

$$r_i = \left| \frac{w_i^{nroi}(C_i^{nroi})}{w_i^{roi}(C_i^{roi})} \right| \tag{4.12}$$

4.6 Experimental Results and Discussion

In order to confirm the validity of the proposed algorithm, we implement it in an image coding framework. Our experiments were conducted with the gray-level Lena image of size 512×512. The MDC system block diagram for this simulation is shown in Figure 4.4. At the polyphase coder, the original input image was first transformed into the frequency domain. Then it was decomposed in the transform domain into S sub-images, i.e., S polyphase components were created. In other words, our image of size 512×512 was first segmented to $\frac{512 \times 512}{S}$ blocks, each of size S pixels. We grouped the image pixels corresponding to the same spatial location in each block to create a polyphase component and S copies

were generated. The polyphase components of each copy were divided by the appropriate psf values corresponding to the position inside each description as shown in Figure 4.4. A set of psf values used in the simulation was computed by the proposed algorithm based on the analytical model-based scheme. Given the total bit rate B, each uncompressed description from MDC codec was then coded with the Said-Pearlman wavelet coder, SPIHT (Set Partitioning In Hierarchical Trees) [69] at a rate $\frac{B}{8}$ and the output bitstream was transmitted as a description. At the receiver, the decoder decided to use the available copy of each polyphase component that has the highest quality for reconstruction.

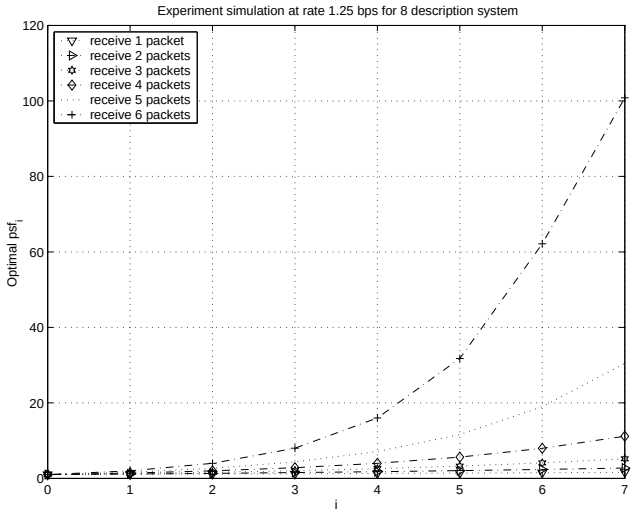


Figure 4.8: Characteristic distribution of psf for the case of 8 descriptions with total bit rate at 1.25 bps

Figure 4.8 shows the characteristics of bit allocation for the varying packet loss rates in terms of the psf assigned in each polyphase component inside a description. psf_i represents the priority scaling factor for polyphase component

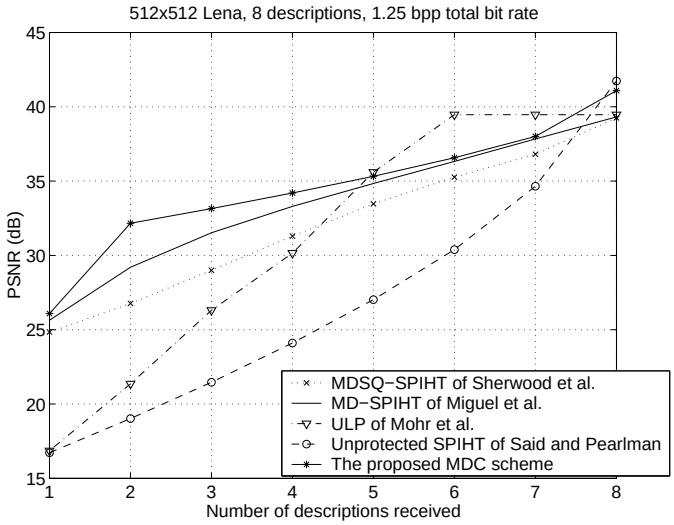


Figure 4.9: Performance comparison between the proposed MDC using the optimal psf values shown in Figure 4.8, Unprotected SPIHT of Said and Pearlman [69], MDSQ-SPIHT of Sherwood *et al.* [72], MD-SPIHT of Miguel *et al.* [45] and ULP of Mohr *et al.* [48] for the case of 8 descriptions with total bit rate at 1.25 bps

i inside each description. As a comparison, when the packet loss rate increases (less number of packets received), the result is that the psf's tends to take similar values. Intuitively, this leads to the equal protection for each polyphase component across all descriptions, as expected. To evaluate the performance of our MDC scheme, we compare the results of our scheme with other MDC works [2] as shown in Figure 4.9. It is clear that our scheme can surpass the unprotected SPIHT [69], MD-SPIHT [45] and MDSQ-SPIHT [72]. However, comparing with ULP [47], our approach is worse when several description have received, because of the advantage of using FEC, although the complexity of our scheme is lower.

[2]The authors would like to thank A. Miguel, A. Mohr and G. Sherwood for providing the performance results.

(a)

(b)

Figure 4.10: Reconstructed image at different packet loss rates for a total of 8 packets: (a) the original Lena image and (b) the reconstructed Lena image when receiving 8 packets

(a)

(b)

Figure 4.11: Reconstructed image at different packet loss rates for a total of 8 packets: the reconstructed Lena images when receiving (a) 6 packets and (b) 4 packets

(a)

(b)

Figure 4.12: Reconstructed image at different packet loss rates for a total of 8 packets: the reconstructed Lena images when receiving (a) 2 packets and (b) 1 packet

To show how graceful degradation is perceptually achieved, we demonstrate the visual qualities of the reconstructed image where different probabilities of loss are applied. Figures 4.10-4.12 illustrate the graceful degradation of subjective quality of the reconstructed images. Furthermore, it is apparent that the results of the proposed MDC system provide robustness even under high packet loss rates, i.e., when only 1 or 2 packets are received.

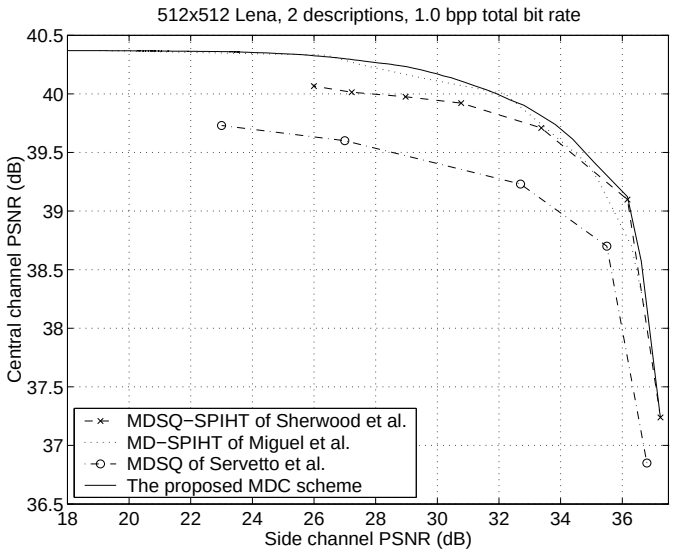


Figure 4.13: PSNR results for our proposed MDC scheme, MDSQ-SPIHT of Sherwood *et al.*, MD-SPIHT of Miguel *et al.*, and MDSQ of Servetto *et al.* for the case of 2 descriptions with total bit rate at 1 bps

The results in Figure 4.13 show the performance comparison of the proposed MDC scheme with the works of Servetto *et al.* [70], Sherwood *et al.* [72] and Miguel *et al.* [45] for the case of 2 descriptions at 1.0 bps. The PSNR results are comparable to the works of [72] and [45] for the low central PSNR region. At high region we improve the performance over the work of [72]. Better performance

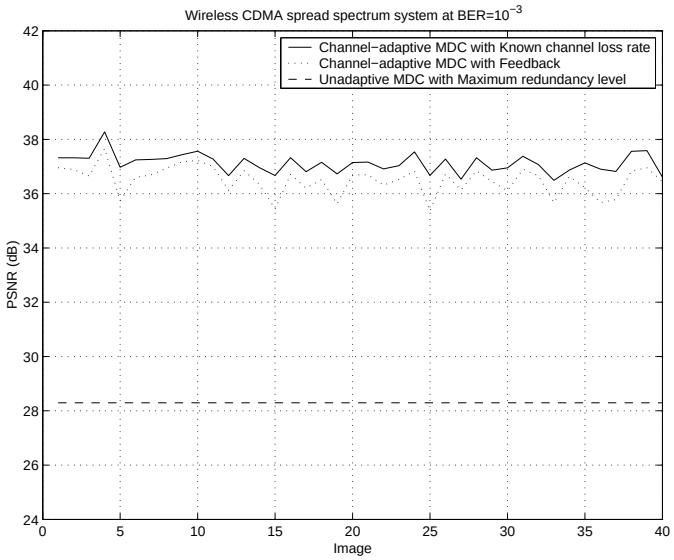


Figure 4.14: PSNR of the gray-scaled Lena image at 1.0 bps in the wireless CDMA spread spectrum system at $BER = 10^{-3}$. The results of the case with known channel loss rate and the case of Equal error protection are shown for benchmarking comparisons.

over the work of [70] can be justified by the better entropy coder by the SPIHT algorithm.

In order to assess the effectiveness in term of adaptivity of the proposed scheme, we show another experiment based on a scenario where a sequence of 40 Lena images is transmitted. In this simulation, each image was coded into 64 descriptions at a total bit rate of 1 bps. One description was transmitted using a packet with 512 bytes payload (an IPv6 node is required to handle 576-byte packets without fragmentation and accounting for packet headers a payload may be as large as 536 bytes [15].)

In this experiment, channel behavior was simulated based on the first-order N-state Markov model to emulate the process of packet losses, which has been shown to be a good approximation in modeling the error process at the packet level [51]. S_i for $i = 0, \ldots, N-1$ denotes the channel states where S_0 represents the good state and all others represent the bad states. In this simulation, we use a 6-state Markov channel model to generate loss patterns. The PSNR results of the end-to-end system for each image are averaged over 50 realizations of the channel patterns. A set of state transition probabilities that is used to emulate a wireless CDMA spread spectrum system at $BER = 10^{-3}$ can be found in [26].

For the proposed scheme, we assume that at the encoder we have no access to the *present* channel behavior, i.e., encoder would not know immediately what the probability of packet lost before encoding process. Instead we assume that there is a feedback channel, which is used only to convey a number of packet losses with a delay of one image interval from the decoder to the encoder. In other words, the packet loss rate used in the optimization process for the image being transmitted is a number of lost packets of the previously transmitted image. Therefore the encoder has *no* knowledge of the statistical model of the channel behavior and can only make use of the *past* packet loss information in the bit allocation algorithm.

To highlight the adaptivity of the proposed scheme, in Figure 4.14, we compare the performances obtained from 3 different scenarios: (i) Channel-adaptive

MDC approach with delayed feedback, (ii) Unadaptive MDC approach with maximum level of redundancy, and (iii) Channel-adaptive MDC approach with known channel loss rate. The first scenario is the proposed MDC scheme with the optimal bit allocation based on the delayed (past) channel information as mentioned earlier, The second scenario is the MDC system *without* adaptation of the redundancy level to the varying network conditions. We set a degree of the redundancy to be maximum, i.e., the worst case scenario. The third scenario serves as a benchmark when the *present* channel information is assumed to be available for optimization at the encoder. This is the unrealistic scenario where the encoder has deterministic knowledge of the future channel conditions. It serves to provide an estimate of the loss in performance due to imperfect channel knowledge in the other cases, and can be thought of as an upper bound on the achievable performance.

Based on our experimental results it is easy to see that the performance, as is to be expected, improves as we increase the information available about the channel conditions. Therefore the performance for the unadaptive MDC scheme is worse than in the case where delayed feedback is available, which in turn has worse performance than the case where future channel loss rates are known. In other words, if there is a mismatch between the probability of packet lost used in the optimization process and the actual probability of packet lost, the performance of the proposed scheme will suffer accordingly. This justifies

the gap of the results between the scenarios (i) and (iii). However the proposed scheme can quickly react to changes in the channel conditions as compared to the unadaptive scheme where it is unchangeably overestimated the packet loss by designing the system based on the worst case packet loss (i.e., maximizing the amount of redundancy).

To validate our idea of combining ROI coding and MDC, we conclude this section by presenting the subjective results in Figure 4.15. when the packet loss rate is 50% and the total bit rate is at 1.25 bps. We illustrate the visual quality when we enable ROI coding feature to have a desired relative distortions at 1 and 4 in Figures 4.15 (a) and (b), respectively. The PSNR of ROI is 35.35 dB and 38.14 in case (a) and (b) respectively while the PSNR of the rest of the image is 35.27 dB and 32.12 in case (a) and (b) respectively. The face region of Lena with ROI coding in Figure 4.16 (a) results in better quality than the one without ROI coding in Figure 4.16 (b). It is worth noting that with ROI coding there is the penalty of lower quality at the background and entire image, but with better quality in the ROI.

4.7 Conclusions

In this work, we investigated the problem of achieving simple adaptation to changing network conditions. We gave an analysis of how a bit allocation technique

(a)

(b)

Figure 4.15: Reconstructed Lena image with 200×200 rectangular-shape ROI located in the middle of the image, i.e., ROI is the area inside the block. for the relative distortions at (a) 1 and (b) 4.

(a)

(b)

Figure 4.16: Zoomed versions of the reconstructed image from (a) Figure 4.15(a) and (b) Figure 4.15(b)

can be used to determine the optimal level of redundancy in each description in an MDC scheme. With the proposed analytical RD model, efficient bit allocation can be achieved based on the divide-and-multiply method with high accuracy as compared to the empirical-based bit allocation method at low complexity. ROI coding was combined with MDC to deliver different levels of protection in different portions of the image. Experimental results showed that the proposed method provided good performance with low complexity by using the proposed analytical model-based bit allocation.

Chapter 5

Dynamic Wavelet Feature-based Watermarking for Copyright Tracking in Digital Movie Distribution System

In the previous chapter we introduced an optimal bit allocation for reliable transmission of data subject to *packet losses*. In this chapter, we consider the need to provide protection to *illegal copying*, when delivering media over a network. We introduce a watermarking system that can be used to uniquely identify the owner and the source of the leak when the media is modified. The main goal is to provide the watermark system to protect the data by distributing the watermark bits into the copyrighted media data [1].

[1] Work presented in this chapter was published in part in [67, 68, 89]

5.1 Introduction

With the recent growth of networked multimedia systems, techniques are needed to prevent (or at least deter) the illegal copying, forgery and distribution of digital audio, images and videos. Many approaches are available for protecting digital data, including encryption, authentication and time stamping. It is also desirable to determine where and by how much a given multimedia file has been changed, when being copied from the original. One way to improve the owner's claim of ownership of an image/video is to place a low-level signal/structure directly into the image/video data. This signal/structure, known as a digital watermark, uniquely identifies the owner and can be easily extracted from the image/video. If the image/video is copied and distributed, the watermark is distributed along with the image/video. This is in contrast to the (easily removed) ownership-information fields allowed by the MPEG-2 syntax. Our goal has been to develop robust digital watermarks for videos to trace any compromised video copy in digital movie distribution systems. The multimedia watermarking algorithm must satisfy the requirement of the non-repudiation of watermarked content in digital distribution systems, which is useful when the source of multimedia contents needs to be known or proofed. One such example is when the content and presentations of that content must be accounted as copyrighted content and the loss of control of that content could lead to monetary loss on the part of the content

provider or the content owner. We are designing a dynamic watermark system to satisfy the following system requirement. First, the watermark should use a key in its watermarking process. This can be used as a primary index key, i.e. there is a large number space from where the key is chosen such that no two keys are likely to be identical if keys are chosen at random. Second, we need to create a unique watermark for each copy of the content. Each watermark will be associated with a different key. Third, we need to create a blank-detect (oblivious) or at least a semi-blank detection watermark. The watermark detection agent should be able to detect the watermark without the original content or with very limited information. Last and most important, we need to create a non-fragile and very robust watermark.

Previously, several digital watermarking methods have been proposed. Cox et al. [12] proposed the spread spectrum based watermark techniques for the video signals, such as FFT/DCT coefficients. Koch, Rindfrey and Zhao [37] also proposed two general methods for watermarking images with DCT techniques. However, the resulting DCT has no relationship to that of the true image and consequently may be likely to cause noticeable artifacts in the image/video and be sensitive to noise. A method for scene-based watermarking of video data was proposed by Swanson, Zhu and Tewfik [77]. In this method, each of a number of frames of a scene of video host data undergoes a temporal wavelet transform, from which blocks are extracted. The blocks undergo perceptual masking in

the frequency domain, such that a watermark is embedded therein. Once the watermark block is taken out of the frequency domain, a spatial mask of the original block is weighted to the watermark block, and added to the original block to obtain the watermarked block. Although this method is somewhat similar to our works, our proposed watermark algorithm is feature-based, in the sense that we alter certain feature values of static frames of wavelet transformations of the video scene sequences. Also, Swanson et al.'s method [77] strictly requires an original video to detect the watermark, while our detection method is semi-oblivious.

In this chapter, we propose, implement and verify novel watermarking embedding and detection algorithms for copyright tracking of digital videos in digital cinema applications. The algorithm is novel in that watermarks created in the wavelet domain of the digital video are unique, dynamic and robust. The watermark patterns are embedded in the special features of wavelet transforms of the original video. Uniqueness of the watermark means, given a digital movie, that the watermark can be identified as a unique label of the movie, or of the scenes/sequences of the movie. Besides, the watermark can be created dynamically according to the time and places of displaying so that the digital watermark can protect the copyright of the multimedia content provider and the further ownership transfers of the media content. Temporal and spatial wavelet decomposition as well as feature-based watermark embedding procedures are deployed

such that the proposed watermarking algorithm is able to effect a compromise tradeoff between visual quality and the robustness. Finally, watermark detection is semi-oblivious. We validated our proposed algorithms with several attacks and the experimental results show that our watermark can survive these attacks up to a level, and that the movie will be unacceptable in terms of quality when the attack is successful.

This chapter is organized as follows: in Section 5.2, we introduce the secure distribution system that our watermark algorithm will be used for. We dictate the specific characteristics of the system architecture such that the watermark framework is clearly defined. We then provide two watermarking algorithms: (i) local-blocking method and (ii) polyphase-based method. We address the embedding and detecting processes of these two methods in Section 5.3. In Section 5.4, experimental results are discussed. The Conclusions of this work are addressed in Section 5.5.

5.2 Secured Digital Media Content Distribution Architecture

More and more digital multimedia data is distributed through public networks. Many approaches are available for protecting digital data; these include encryption, authentication, time stamping and watermarking. Most existing watermark

schemes in distributed systems depend on a trusted third party (TTP) to verify the authentication of the watermark system. The secure delivery of images over open networks proposed by Augot et al [4] may encounter situations such that a "trusted third party" can not be found that can be trusted by both parties. Other watermarking systems that we are aware of that aim to accomplish the same end goal must employ the services of a "trusted third party" to put watermarking keys in escrow to be presented upon demand if there is a dispute.

We propose a mechanism, which does not need "trusted third party", every watermark is non-repudiation watermark, and can be used to identify the source of the watermark. Although there are many uses for this technology, the use for which it was developed is multimedia content distribution forensic analysis. In these cases, the multimedia content must be kept secret and not distributed by unauthorized agents. Should the content "leak" and become uncontrolled, it is desirable to locate the source of the leak so that appropriate action can be taken (punitive damages sought and security tightened for instance.) This multimedia watermarking mechanism allows content to be linked to the end user or distributor of the content, whichever is the responsible party in a way that corrective action (legal or technical) may be taken with confidence.

To show how the digital content will be safely distributed, we present a distribution model as shown in Figure 5.1. The content provider passes the valuable digital content to the content distributor to be distributed to all eligible clients,

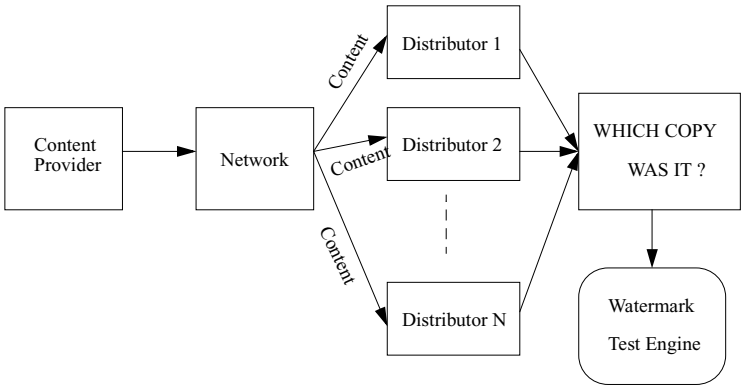


Figure 5.1: Digital content distribution model with secure copy monitoring

in this case, the clients are also distributors. All clients are required to put the watermarks into the digital content according to the proposed non-repudiation watermark schema so that content provider can trace the source of the leak if the copy is leaked. The novelty of the distribution model and non-repudiation watermark schema is that it allows the reliable and non-repudiable watermark to fulfill the needs and trust of content provider and content consumers/presenters. This distribution model does not implement any protections to ensure that the watermark is applied properly, it is however assumed that both parties (the content provider and the distributor) willingly agree to follow the procedure as outlined. In the case where the distributor or the provider wishes to "cheat" the other by circumventing the watermarking procedure, other measures must be taken to ensure that this is not done [61]. Also, watermark attacks must be addressed and considered in designing suitable watermark algorithm [68] for the system.

The non-repudiation watermark schema for digital multimedia distribution is depicted in Figure 5.2. This schema requires the use of public and private key encryption algorithms and assumes the participation of one content provider and at least one content distributor. Both the content provider and the distributor have their own private key that they do not share. This key is central to the identification of content watermarked by the distributors. First the content provider sends a file with the content to the distributor. This may be done in a variety of ways including but, not limited to, transmission of the content through a data network and distribution of the content on physical media (for instance CDROM's or DVD's). Once the content has been sent to the distributor, the distributor must contact the provider. This contact must be authenticated using "strong" authentication techniques. The exchange must be protected by "strong" encryption techniques. After authenticating with one another, the provider provides a random number to the distributor. The size of this random number is dictated by the watermark style to be used. Multiple watermark mechanisms may be used with this technology. The distributor generates a public and private key pair for the application of this watermark, which we will call the watermark pair. The distributor uses the watermark pair private key to encrypt the random number passed to it by the provider. This encrypted number will be the watermark key. The distributor then watermarks the content with this key. After watermarking (or before, depending on the details of the watermark process), the distributor

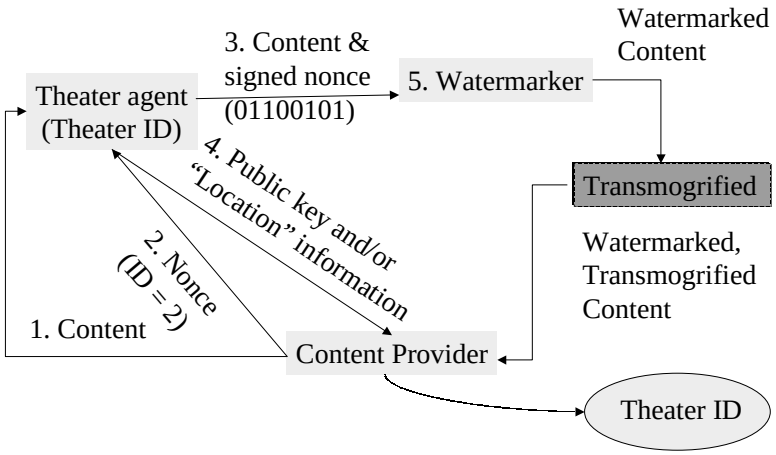


Figure 5.2: Non-repudiate watermark scheme for digital movie distribution

will transmit to the provider the public key of the watermark pair signed by its private key (not the watermark pair private key, the distributor private key) along with information that will allow the provider to obtain the watermark key given the watermarked content. We will label this information "location information". However, it may not designate a location in the traditional sense of the word and will depend on the watermark technique selected. The provider archives this information so that the watermark may be detected later.

5.3 Watermark Algorithms

We are proposing a scene-adaptive feature-based watermarking algorithm and method in video's 3-D wavelet domain. Figure 5.3 shows the general watermarking embedding procedure. First, video is segmented into scenes. Digital movies

are composed of many scenes, each of which has a sequence of similar images. We use the methods proposed by Zhou et al. [90] for scene change detection and separate the movie into multiple scenes. That is, the histogram of each frame is compared to that for its adjacent frames to detect the boundary of a scene. Each scene will be separately cast with a watermark pattern. Inside each scene, we develop an algorithm to hide the data in such a way that the comparison between frames can not lead attackers to detect, or even worse remove, the embedded watermarks. Since the characteristics of each scene are totally different from those of neighboring scenes, casting different watermark patterns with an independent frame in each scene would easily result in noticeable artifacts. Therefore, we propose to embed one type of watermark per scene, or part of a watermark sequence per scene. To avoid averaging and collusion attacks, for example, if different watermarks were embedded in the frames inside the same scene, they could be easily removed by comparing frames or simply averaging the frame with its neighbors to construct new frames without visually noticeable degradation. We then propose to embed a single feature-based watermark pattern for all the frames inside a scene on a scene-by-scene basis. To avoid collusion at the frames near the scene boundary, we need to cast different watermarks in neighboring scenes.

In our proposed watermarking framework, the main focuses are (i) to determine *where* in a video sequence a watermark can be hidden so that an optimal tradeoff between robustness and visual quality can be achieved, and (ii) to learn

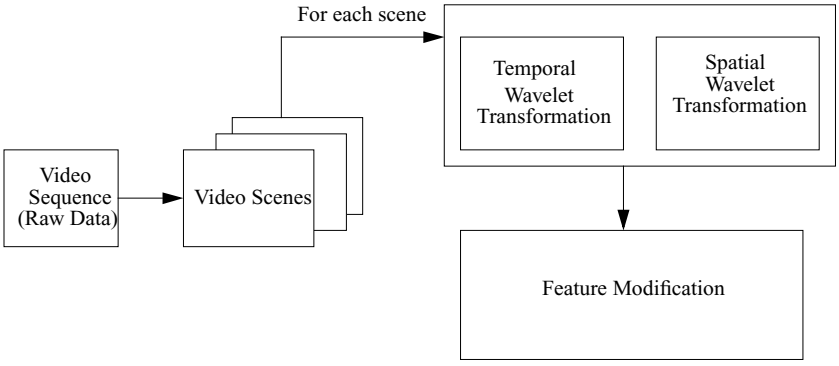
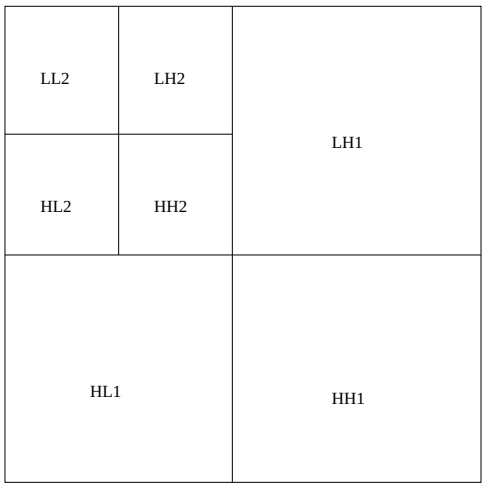
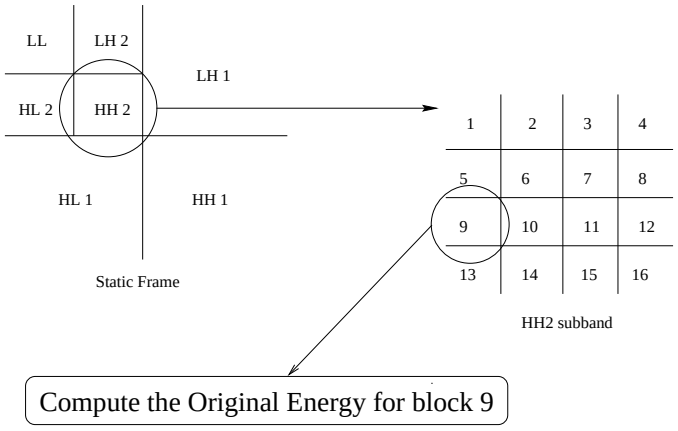


Figure 5.3: Digital video watermark embedding procedure

how to cast a watermark pattern into the selected places in a video sequence so that the given tradeoff in (i) is satisfied and the watermark detection algorithm can be performed *without* the original video sequence. Since wavelet transformation has shown promising ability in classifying video streams into multiple bands with different characteristics, we introduce the watermark in the wavelet domain. We first apply a temporal wavelet decomposition to the original video sequence to obtain: (i) static frames and (ii) dynamic frames. Static frames basically are obtained by applying a wavelet low-pass filter along the temporal domain and subsampling the filtered frames by 2. Dynamic frames are obtained in the same way but using a high-pass filter instead. For instance, After applying 4-level decomposition to a 144-frame video sequence it will result in 9 static frames and 135 dynamic frames. Furthermore, we apply the spatial wavelet decomposition to each *static* frame. For example, 2-levels wavelet decomposition will result in



(a)



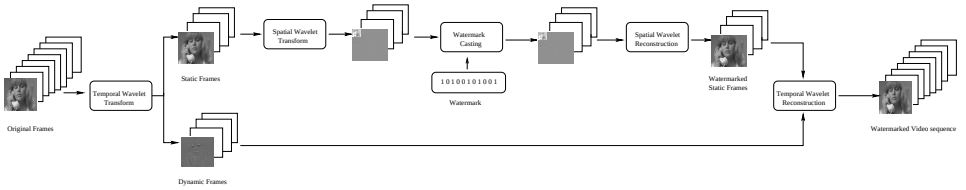
(b)

Figure 5.4: (a) Multiple resolution bands after performing spatial wavelet decomposition and (b) watermark casting

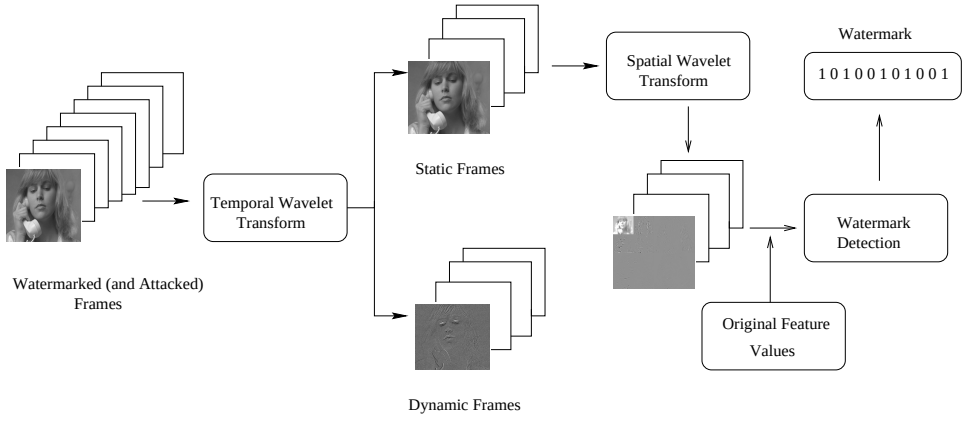
7 subbands, i.e., LL2, LH2, HL2, HH2, LH1, HL1 and HH1, as shown in Figure 5.4(a). The subband LL2 represents the approximation of the static frame, which contains the most important data. The other bands contain high frequency information such as the edge information of each static frame. In summary, in the *temporal domain*, we choose to cast the watermark in *static* frames so that the watermark can be spread all over every frame in the scene. In the *spatial domain*, we decide to embed the watermark pattern on the *middle frequency* bands, e.g., HL2, HH2 and LH2, to establish a tradeoff between the robustness of watermarks and the visual quality of the video sequence. Figure 5.5(a) shows the watermark casting procedure.

Next we need to determine *how* the watermark bits will be cast into the preselected middle frequency bands of static frames. In this work, we propose a feature-based watermark-casting algorithm. Basically, it can be any feature in the selected subband. Here, we use a simple feature such as the energy of the wavelet coefficients blocks.

As shown in Figure 5.4(b), we start by locally separating the coefficients nearby into multiple blocks. We use the classic embedding procedure of adding noises into images [87] for the watermark insertion. But different from this previous work [87], we change the value of the selected features, the energy of the wavelet coefficients of selected subbands. For the i^{th} selected subband, the energy of each block, E^o_{ij} for the j^{th} block, is modified to be the watermarked energy



(a)



(b)

Figure 5.5: (a) Block diagram of the proposed watermark casting technique when the watermarks are embedded into a single scene of video sequence of 8 frames and (b) block diagram of the proposed watermark detection for a single video sequence

E_{ij}^{w} , where α is watermark strength and the b_j is the j^{th} watermark bit, which is either 0 or 1, i.e., $E_{ij}^{w} = E_{ij}^{o}(1 + \alpha b_j) = E_{ij}^{o} + E_{ij}^{o}\alpha b_j$. At the i^{th} selected subband, since the energy is the summation of the squared wavelet coefficients, $E_{ij}^{o} = \sum_{k=1}^{N_i} |x_{ijk}^{o}|^2$ and $E_{ij}^{w} = \sum_{k=1}^{N_i} |x_{ijk}^{w}|^2$. N_i is the total number of wavelet transformed coefficients in one block that belongs to the i^{th} selected subband. x_{ijk}^{o} and x_{ijk}^{w}, of the k^{th} selected subband, are the k^{th} wavelet coefficients in the j^{th} original block and the j^{th} watermarked block, respectively. Then the original coefficient can be modified linearly to obtain a watermarked coefficient , i.e., $x_{ijk}^{w} = (\sqrt{1 + \alpha b_j})x_{ijk}^{o}$. We only cast the watermark bits into the middle frequency bands of LH1, HL1 and HH2. After we have cast all watermark bits into all the static frames of the scene, we then reconstruct the movie into raw data domain by spatial and temporal inverse-wavelet transformation.

To detect the cast watermark pattern, we repeat some of the processes in the watermark casting procedure i.e., scene change detection, temporal and spatial wavelet decomposition, as shown in Figure 5.5(b). Watermark key detection takes place in the following way. The content provider obtains a copy of the watermarked content, and then iteratively attempts to extract the watermark key using the location information provided by each distributor. After a candidate watermark key is extracted, an attempt is made to verify the key by decrypting the key with the distributor's public key and comparing it with the random number, which was sent to the distributor at the start of the initial exchange by

the content provider. If the numbers match, the copyright of the content has been successfully identified. Otherwise, provider continues with the key of the next distributor. At the i^{th} selected subband, the gap in energy between casting or not casting a watermark bit is the range between 0 and $E^o_{ij}\alpha$. The watermark bits are detected according to Equation 5.1.

$$\begin{aligned} E^w_{ij} - E^o_{ij} \geq \frac{E^o_{ij}\alpha}{\beta} \quad &: \textit{embedded watermark bit} \;=\; 1 \\ E^w_{ij} - E^o_{ij} < \frac{E^o_{ij}\alpha}{\beta} \quad &: \textit{embedded watermark bit} \;=\; 0 \end{aligned} \tag{5.1}$$

where the threshold $\frac{E^o_{ij}\alpha}{\beta}$ is somewhere between the two ends, and β is the parameter, which we can adjust to an appropriate threshold for watermark detection. In the experiment section, we will use $\beta = 2$. We perform the above process repeatedly until all the watermark bits in every static frame have been detected. Note that our proposed watermark algorithm does *not* need to use the original movie to detect the watermark bit, but only depends upon passing the original feature values to the detector, i.e., E^o_{ij} and α. This is very useful when a third trusted party does not exist in the security-system point of view.

After a candidate watermark key is extracted, iterative attempts are made to verify the distributor's identification by decrypting the key with the distributors' public key (provided at the watermarking time in schema of Figure 5.2) and then

compared with the nonce in Figure 5.2 (given to the distributor at the start of the initial exchange.) If the decrypted watermark key matches the nonce, the content source has been successfully identified, if not, the producer goes on to the next distributor.

5.4 Experimental Results and Discussion

We simulated the watermark on the video test sequence Suzie and Akiyo, which each has 144 frames, and assumed the video was a scene of a movie. Each frame has a size of 144x176. We then applied 4-level temporal and 2-level spatial wavelet decompositions. We categorize our experiment results into a few classes and present them in the following subsections based on the visual quality and the robustness against various attacks.

If we put 8 watermark bits in each static frame, the watermark payload can be up to 72 bits with all the 9 static frames included. This is sufficient for the digital cinema applications. We used the average Peak Signal to Noise Ratio (PSNR) as an objective performance measurement of the visual quality after the watermark embedding. We changed the strength of the watermark based on the parameter α and plotted the average PSNR for both the Suzie, as shown in Table 5.1. As expected, the higher the strength of the watermark casting, the worse the video quality, showing the tradeoff between robustness and visual quality. Thus we

(a)

(b)

(c)

Figure 5.6: Comparison of visual quality between (a) original Suzie frame and (b) watermarked Suzie frame when (c) represents the watermark embedded

α	1.0	1.5	2.0	2.5
Visual quality (dB)	50.36	49.05	47.90	46.90
MPEG compression (kbps)	450	395	335	295
Temporal cropping	37.50 %	40.28 %	46.71 %	59.72 %
Temporal dropping	66.67 %	75.00 %	80.00 %	80.00 %
Temporal Averaging	YES	YES	YES	YES
Spatial cropping	35.23 %	55.68 %	65.91 %	68.18 %
Spatial dropping	NO	25.00 %	25.00 %	25.00 %
Spatial Averaging	NO	YES	YES	YES

Table 5.1: Experimental results of QCIF-formatted Suzie sequence shows the trade-off between visual quality and robustness at different watermark strengths, α. Percentage of temporal cropping/dropping is computed as a ratio of the cropped/dropped *frames* and the total number of the coded frames in the sequence. Percentage of spatial cropping/dropping is computed as a ratio of the cropped/dropped *pixels* per frame and the total number of the coded pixels per frame. These percentages indicate the maximum degree of corrupted data allowed before the watermark detection fails to correctly extract the embedded information. For a given watermark strength α, YES/NO indicates the success/failure of watermark detection.

have to select an appropriate watermark strength, α, so as to compromise the visual requirements excessively. To validate the perceptual quality, Figure 5.6 shows the visual differences between the original and watermarked images, and no visual difference of the two images can be noticed.

The upper bound of the distortion can be approximated: if all the watermark bits are 1's, then the energy of each block has to be increased. This is the case of maximum degree of modification. All the wavelet coefficients in the selected subbands (i.e., LH1, HL1 and HH2) will be modified to increase in the order of $x^w_{ijk} - x^o_{ijk} = (\sqrt{1+\alpha b_j} - 1)x^o_{ijk} \leq (\sqrt{1+\alpha} - 1)x^o_{ijk}$. Thus, the average Mean Square Error (MSE) can be approximated to be

$$
\begin{aligned}
MSE &= \frac{\sum_{i=0}^{N_b}\sum_{j=0}^{N_w}\sum_{k=0}^{N_i}|x_{ijk}^{w}-x_{ijk}^{o}|^2}{N_b \times N_w \times \sum_{i=0}^{N_b} N_i} \\
&\leq (\sqrt{1+\alpha}-1)^2 \frac{\sum_{i=0}^{N_b}\sum_{j=0}^{N_w}\sum_{k=0}^{N_i}|x_{ijk}^{o}|^2}{N} \\
&= (\sqrt{1+\alpha}-1)^2 E^o \qquad (5.2)
\end{aligned}
$$

where N_b is the number of selected subbands, N_w is the number of watermark bits, N_i is the total number of wavelet coefficients for one block of the i^{th} selected subband, N is the total number of wavelet coefficients in all selected subbands, and, finally, E_o is the average original energy of all selected subbands. From Equation (5.2), it is obvious that if the average original energies of all the selected subbands are given, we can determine approximation of the objective quality, MSE. This approximate relationship between the MSE and the watermark strength, α, will be highly useful as a tradeoff between the quality of the watermarked movie and the robustness of the watermarks. In this way, different watermark strengths can be determined based on the particular characteristics of the embedding scene.

To validate the robustness of the proposed watermarking method, we started by applying an MPEG re-compression attack to the watermarked testing sequences to different bit rates. Table 5.1 shows that the embedded watermarks can survive at 2/3 of the original video bit rates after re-compression to lower bit rates.

Frame dropping, or temporally frame subsampling, can occur when frame rates are changed. As shown in Table 5.1, we can detect all watermark bits correctly when the frames are dropped up to 80% (i.e. dropping 4 out of 5 frames). Frame cropping is to cut video sequences shorter by deleting the frames in the beginning/end. As shown in Table 5.1, our algorithm can survive at very high cropping rate. Besides, to detect a watermark after frame sampling/dropping attacks, we substituted the missed frames with the average of the frames available. All watermark bit can be detected correctly when subsampling/dropping rate is up to 80%, which means that our proposed algorithm can also tolerate the frame averaging attack.

Spatial cropping is clipping image row(s) or column(s). For instance, a wide-screen movie may need spatial cropping to fit on normal TV screens. As shown in Table 5.1, our algorithm performs very well even when only a small percentage of the original image is retained. Spatial dropping or subsampling is the selective removal of rows and/or columns of an image. Our watermark survives when only 1/4 of the original image is retained. High resolution movies are more vulnerable to spatial dropping or subsampling attacks. It is worth nothing that the detection against cropping/sub-sampling requires range searching and shifting.

5.5 Conclusions

Initial experiments show that our proposed technique has good potential for copyright tracking in digital cinema applications. By implementing the wavelet-transformation in hardware, the computation complexity of the 3D wavelet transformation is expected to be decreased. In the future, we plan to extend our proposed semi-oblivious watermarking technique to oblivious detection, and test these algorithms with more attacks, especially geometrical attacks.

Chapter 6

Future Work

In this chapter we propose preliminary ideas for possible extensions of our approaches.

6.1 ROI coding for JPEG 2000

Besides JPEG2000 [6] has become one of the most recent powerful tool in RD-optimized image compression, it also provides several useful functionality in many diverse application areas. Region of interest (ROI) coding is one of the popular feature sets where the variation of fidelity can be progressively varied in spatial domain. In chapter 3, we have already addressed the analytical bit allocation algorithm for ROI coding for SPIHT codec. However SPIHT is *not* the image compression standard that is currently most widely used. Adjustments of the proposed model-based bit allocation algorithm are needed to be able to deliver

good performance for the JPEG2000 system, which is emerging as the most effective international image coding standard.

Since the scaling-based method is performed immediately before quantization and entropy coding, it can be easily applied to JPEG2000 system. That is, after wavelet-transform coding at JPEG2000 encoder, we can divide the wavelet coefficients associated with non-ROI by the psf value and then multiply the same value back at the decoder immediately before the inverse wavelet-transform is performed.

In the proposed RD model from Chapter 3, recall that the total bitrate is converted to the total number of significant coefficients by,

$$M_b = \frac{NR_b}{\mathbf{6.6}} \tag{6.1}$$

where 6.6 is the number of bits per significant coefficient obtained by performing experiments for the SPIHT codec over a hundred of test images at different bitrates [42, 41]. The RD model thus has to be slightly modified to find what a correct parameter would be for JPEG 2000 by collecting the data from the testing images. Once the number is found, the proposed model can be directly used and our analytical bit allocation algorithm can efficiently work under JPEG2000 system.

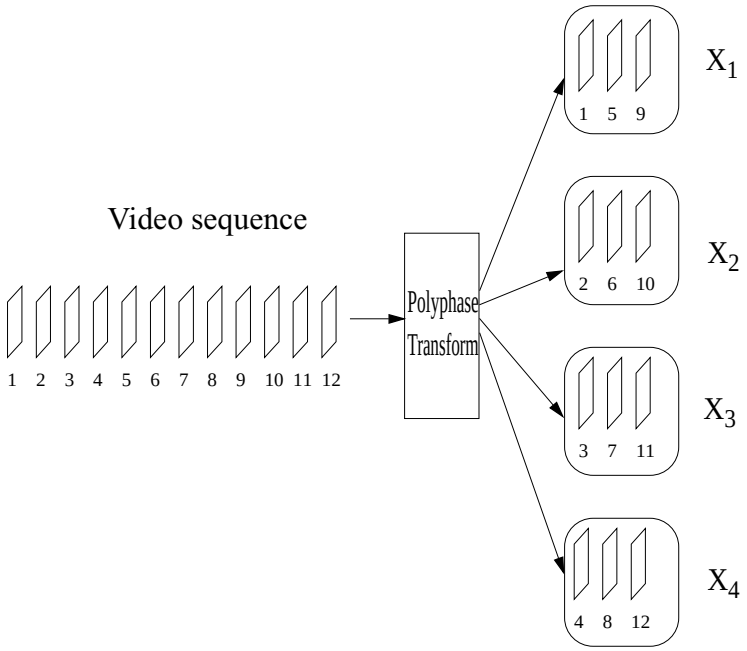


Figure 6.1: An example of polyphase transform for video MDC when an original video sequence, which is assumed to have a total of 12 frames, is segmented into 4 subsequences. Each subsequence is composed of 3 frames and represent a polyphase component.

6.2 Polyphase-based MDC for Video Coding

In Chapter 4 we have proposed a simple MDC approach for image source. However it is possible to extend the proposed idea to the video source. More specifically, the polyphase transform can be used to split the sequence of video frames into a number of separate subsequences as shown in Figure 6.1. Each subsequence is considered as a polyphase component and will be packetized into a description. Each description will be coded using any existing video encoder, e.g., H.26x or MPEG-x. The decoder operates by gathering the available information for each

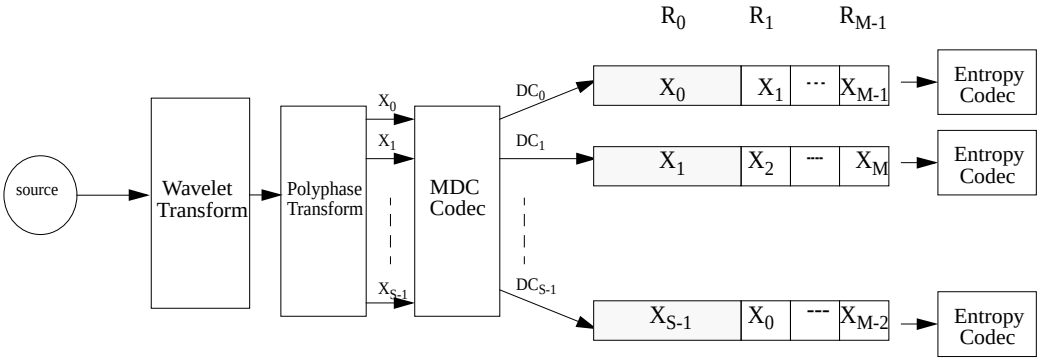


Figure 6.2: Block diagram when psf represents the quantization parameter for each frame

subsequence. The it selects for each subsequence its highest quality copy to be used in the decoding process.

The bitrate for each frame can be adjusted by choosing the appropriate quantization parameter, leading to the video MDC system as shown in Figure 6.2. Our goal is to determine the best set of quantization parameters (i.e., $\{psf_0, \ldots, psf_{S-1}\}$) given the probability of packet loss. To determine the optimal bit allocation (i.e., the best set of quantization parameters), several approaches can be used to reduce the complexity as compared to the traditional exhaustive search technique. The message-passing-based bit allocation algorithm for temporally dependent coding introduced in Chapter 2 can be used to reduce the complexity and memory requirements. The approach proposed by Ortega *et al.* [56] can also be incorporated to provide the optimal solution.

The advantage of the proposed video MDC scheme over most of the existing video MDC approaches [35, 38] is the absence of drift effect. More specifically,

we use separate motion estimation/compensation processes for each description to avoid the mismatch in prediction loop. Unlike the work proposed by Apostolopoulos [2, 1], which also eliminate the drift effect, our proposed approach can be viewed as a more general case since we can choose to transmit more than one polyphase component in one description. If the channel conditions are bad, our system can select to send more than one polyphase component by optimally distributing the bitrate budget for each description among all polyphase component. However if we operate in an error-free environment, we can switch off the MDC system to provide no redundancy, i.e., transmit only one polyphase component for one description. Clearly Apostolopoulos's approach [2, 1] is a special case of our approach when operating in the error-free channel, i.e., only one polyphase component in a description. In other words, our approach provides additional flexibility in terms of choosing the amount of redundancy.

6.3 Oblivious Watermarking

With the proposed watermarking algorithm, we require to have the original energy for the watermark detection. In particular, each original energy has to match the associated scene such that the detection can be correctly performed. That means if the attacker switches the order of the scene, the watermark detection will fail.

Therefore it would be more secure to modify the proposed watermarking algorithm such that none of the information from the original video sequence is required in the detection process. With the well-established watermarking algorithm we proposed in Chapter 5 the location to cast the watermark remains unchanged, middle frequency subbands of the static frames. However, an alternative watermark insertion algorithm is needed. Energy seems to be the simplest choice to modify but other features are possible too. In other words, beside the energy modification, other features of wavelet coefficient can be considered as the good candidates for watermarking. The challenging question would be how to choose a feature that (i) can provide the robustness against the attacks (confidential security) and (ii) can lead to the least noticeable artifacts in the host data (perceptuality). Also that feature would not be needed to be available in the detection process (oblivious detection). Furthermore, a simple 3D visual model can be used to determine the appropriate watermark strength to improve the tradeoff between fidelity and robustness.

Reference List

[1] J. Apostolopoulos. Reliable video communication over lossy packet networks using multiple state encoding and path diversity. *VCIP 2001*, Jan. 2001.

[2] J. Apostolopoulos and S.J. Wee. Unbalanced multiple description video communication using path diversity. *IEEE 2001 International Conference on Image Processing, ICIP 2001*, Oct. 2001.

[3] E. Atsumi and N. Farvardin. Lossy/lossless region-of-interest image coding based on set partitioning in hierarchical trees. *Proc. IEEE International Conference on Image Processing (ICIP 98, pp. 87-91)*, Oct. 1998.

[4] D. Augot, J-M. Boucueau, J.F. Delaigle, C. Fontaine, and E. Goray. Secure delivery of images over open networks. *Proceedings of the IEEE,Vol. 87, No. 7*, July 1999.

[5] S. Benedetto, D. Divsalar, G. Montorsi, and F. Pollara. Soft-input soft-output modules for the construction and distributed iterative decoding of code networks. *European Trans. Comm.*, Mar. 1998.

[6] M. Boliek. JPEG 2000 Part I Final Commitee Draft V. *ISO/IEC JTC1/SC29/WG1 N1646R*, Mar 2000.

[7] K.M. Chugg, A. Anastasopoulos, and X. Chen. *Iterative Detection*. Kluwer Academic Publishers, 2001.

[8] K.M. Chugg and X. Chen. Efficient architectures for soft-output algorithms. *Proc. Intl. Conf. Comm., paper S04P4*, Jun. 1998.

[9] K.M. Chugg, X. Chen, A. Ortega, and C.-W. Chang. An iterative algorithm for two-dimensional digital least metric problems with applications to digital image compression. *IEEE 1998 International Conference on Image Processing (ICIP 1998), Chicago, Illinois*, Oct. 1998.

[10] D. Comas, R. Singh, and A. Ortega. Rate-distortion optimization in a robust video transmission based on unbalanced multiple description coding. *2001 Workshop on Multimedia Signal Processing*, 2001.

[11] T. M. Cover and J. A. Thomas. *Elements of information theory.* Wiley series in telecommunications, Wiley, 1991.

[12] I. J. Cox, J. Kilian, F. T. Leighton, and T. Shamoon. Secure spread spectrum watermarking for multimedia. *IEEE Transactions on Image Processing, Vol. 6, No. 12*, Dec 1997.

[13] C. Cristopoulos, A. N. Skodras, and T. Ebrahimi. JPEG 2000 still image coding system: An overview. *IEEE Trans. Consumer Electron., Vol. 46, pp. 1103-1127*, Nov 2000.

[14] D. Santa Cruz, T. Ebrahimi, M. Larsson, J. Askelof, and c. Cristopoulos. Region of interest coding in JPEG2000 for interative client/xerver applications. *Proc. of the IEEE Workshop on multimedia signal processing (MMSP'99), pp. 389-394*, Sept. 1999.

[15] S. Deering and R. Hinden. Internet protocol, version 6 (ipv6) specification, network working group request for comments 1883. Dec. 1995.

[16] S. Dolinar, A. Kiely, M. Klimesh, R. Manduchi, A. Ortega, S. Lee, H. Xie P. Sagetong, J. Harel G. Chinn, S. Shambayati, and M. Vida. Region-of-interest data compression with prioritized buffer management. *2001 Earth Science Technology Conference (ESTC 2001), College Park, MD*, Aug. 2001.

[17] S. Dolinar, A. Kiely, M. Klimesh, R. Manduchi, A. Ortega, S. Lee, H. Xie P. Sagetong, J. Harel G. Chinn, S. Shambayati, and M. Vida. Region-of-interest data compression with prioritized buffer management (II). *2002 Earth Science Technology Conference (ESTC 2002)*, Aug. 2002.

[18] S. Dolinar, A. Kiely, M. Klimesh, R. Manduchi, A. Ortega, S. Lee, H. Xie P. Sagetong, J. Harel G. Chinn, S. Shambayati, and M. Vida. Region-of-interest data compression with prioritized buffer management (III). *2003 Earth Science Technology Conference (ESTC 2003)*, Aug. 2003.

[19] A. A. El-Gamal and T. M. Cover. Achievable rates for multiple descriptions. *IEEE Trans. Information theory, vol. IT-28, no. 6, pp. 851-857*, Nov. 1982.

[20] H. Everett. Generalized Lagrange multiplier method for solving problems of optimum allocation of resources. *Operations Research, vol. 11, pp. 399-417*, 1963.

[21] FIPS 186. *Data Encryption Standard.* 1977.

[22] M. Fumagalli, P. Sagetong, and A. Ortega. Estimation of erased data in a h.263 coded stream by using unbalanced multiple description coding. *IEEE International Conference on Multimedia and Expo (ICME)*, 2003.

[23] V. K. Goyal. Multiple description coding: compression meets the network. *IEEE Signal Processing Mag., vol. 18, pp. 74-93*, Sept. 2001.

[24] V. K. Goyal, J. Kovacevic, R. Arean, and M. Vetterli. Multiple description transform coding of images. *Proceeding of ICIP-98*, Oct. 1998.

[25] V. Hardman, A. Sasse, M. Handley, and A. Wason. Reliable audio for use over the internet. In *Proc. INNET*, 1995.

[26] C. Y. Hsu, A. Ortega, and M. Khansari. Rate control for robust video transmission over burst-error wireless channels. *IEEE JSAC special issue on multimedia network radios, vol. 17, No. 5, pp. 756-773*, May 1999.

[27] C.-Y. Hsu, A. Ortega, and A.R. Reibman. Joint selection of source and channel rate for vbr video transmisson under atm policing constraints. *IEEE Journal on Sel. Areas in Comm., pp. 1016-1028*, Aug. 1997.

[28] ISO/IEC 11172. Information technology - coding of moving pictures and associated audio for digital storage media at up to about 1.5 mbit/s. 1993.

[29] ISO/IEC 13818. Information technology - generic coding of moving pictures and associated audio information. 1996.

[30] ITU-T Recommendation H.261. Video codec for audiovisual services at px64 kbits/s. 1993.

[31] ITU-T Recommendation H.263. Video coding for low bitrate communication. 1996.

[32] W. Jiang and A. Ortega. Multiple description coding via polyphase transform and selective quantization. In *Proceedings of VCIP'99*, pages 768–778, San Jose, CA, Jan. 1999.

[33] Joint Video Team of ITU-T and ISO/IEC JTC 1. Draft ITU-T Recommendation and Final Draft International Standard of Joint Video Specification (ITU-T Rec. H.264 ISO/IEC 14496-10 AVC). *Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG, JVT-G050.*

[34] L. Ke and M.W. Marcellin. Near-lossless image compression: Minimum-entropy, constrained-error DPCM. *IEEE Transaction on Image Processing, Volumn 7, Issue 2, pp. 225-228*, Feb. 1998.

[35] C.-S. Kim and S.-U. Lee. Multiple description motion coding algorithm for robust video transmission. *Proc. IEEE Int. Symp. Circuit and Systems 2000, Lausnne, Switzerland, vol. 4, pp. 717-720*, May 2000.

[36] N. Koblitz. *A course in number theory and cryptography.* Springer, 1994.

[37] E. Koch, J. Rindfrey, and J. Zhao. Copyright protection for multimedia data. *In Proc. Int. Conf. Digital Media and Electronic Publishing*, 1994.

[38] K.-W. Lee, R. Puri, T. Kim, K. Ramchandran, and V. Bharghavan. An integrated source coding and congestion control framework for video streaming in the internet. *Proc. IEEE INFOCOM, Tel Aviv, Israel, vol. 2, pp. 747-756*, Mar. 2000.

[39] K. Lengwehasatit. Complexity-distortion tradeoffs in image and video compression. *Ph.D. Thesis*, Dec. 1999.

[40] S. Lin and Jr. D. J. Costello. *Error control coding: Fundamentals and applications.* Prentice-Hall, 1983.

[41] S. Mallat. *A Wavelet Tour of Signal Processing.* Academic Press, New York, 1998.

[42] S. Mallat and F. Falzon. Analysis of low bit rate image transform coding. *IEEE Trans. on Signal Processing, vol. 46, no. 4, pp. 1027-1042*, April 1998.

[43] S. McCanne, M. Vetterli, and V. Jacobson. Low complexity video coding for receiver driven layered multicast. *IEEE Journals on Selected Areas in Communications, pp. 983-1001*, Aug. 1997.

[44] A. Miguel and E. Riskin. Protection of region of interest against data loss in a generalized multiple description framework. *Proc. of the Data Compression Conference (DCC)*, 2000.

[45] A. C. Miguel, A. E. Mohr, and E. A. Riskin. SPIHT for generalized multiple description coding. In *Proc. of ICIP-99*, Kobe, Japan, Oct 1999.

[46] C. Miller, B.R. Hunt, M.A. Neifeld, and M.W. Marcellin. Binary image reconstruction via 2-d viterbi search. *IEEE 1997 International Conference on Image Processing (ICIP 1997), Santa Babara, CA*, Oct. 1997.

[47] A. E. Mohr, E. A. Riskin, and R. Ladner. Generalized multiple description coding through unequal forward error correction. In *Proc. of ICIP-99*, Kobe, Japan, Oct 1999.

[48] A.E. Mohr, E.A. Riskin, and R.E. Ladner. Generalized multiple description coding through unequal loss protection. *CSTV'99*, Sept. 1999.

[49] D.L. Neuhoff, T.N. Pappas, and N. Seshadri. One-dimensional least-squares model-based halftoning. *Proc. ICASSP, San Francisco, CA, pp/ 189-192*, Mar. 1992.

[50] D. Nister and C. Christopoulos. Lossless region of interest with a naturally progressive still image coding algorithm. *Proc. IEEE International Conference on Image Processing (ICIP 98, pp. 856-860)*, Oct 1998.

[51] A. Ortega and M. Khansari. Rate control for video coding over variable bit rate channels with applications to wireless transmission. *Proc. of Intl. Conf. on Image Proc., ICIP'95, Washington D.C.*, Oct. 1995.

[52] A. Ortega and K. Ramchandran. Rate-distortion methods for image and video compression. *IEEE Signal Processing Magazine, pp. 23-50*, Nov. 1998.

[53] W.B. Pennebaker and J.L. Mitchell. *JPEG Still Image Data Compression Standard*. Van Nostrand Reinhold, New York, 1993.

[54] B. Prakash, K.N. Ramakrishnan, A.G. Suresh, S. Chow, and T.W.P. Fetal lung maturity analysis using ultrasound image features. *IEEE trans. on Information technology in biomedicine, vol. 6, issue 1, pp. 38-45*, Mar. 2002.

[55] R. Puri and K. Ramchandran. Multiple description source coding through forward error correction codes. *Proceedings of the 33'rd Asilomar conf. on Signals, systems, and computers, Pacific Grove, CA*, Oct. 1999.

[56] K. Ramchandran, A. Ortega, and M. Vetterli. Bit allocation for dependent quantization with applications to MPEG video coders. *Proceeding of ICASSP'93*, April 1993.

[57] K. Ramchandran, A. Ortega, and M. Vetterli. Bit allocation for dependent quantization with applications to multiresolution and MPEG video coders. *IEEE Trans. on Image Proc., vol. 3, pp. 533-545*, Sept. 1994.

[58] R. Rejaie, M. Handley, and D. Estrin. Rap: And end to end rate based congestion control mechanism for real-tiume streams in the internet. *Proc. of INFOCOM*, 1999.

[59] R.L. Rivest, A. Shamir, and L. Adleman. A method for obtaining signatures and public-key cryptosystems. *ACM, vol. 21, no. 2, pp. 120-126*, Feb. 1978.

[60] L. Rizzo. Effective erasure codes for reliable computer communication protocols. *ACM Computer Comm. Rev., vol.27, pp. 24-36*, Apr. 1997.

[61] T. Rockwood, W. Zhou, B. Ryu, and Y. Zhang. Secure systems and methods for digital cinema distribution. *Techinical Report, Information Sciences Lab, HRL Laboratories, LLC.*, June 2001.

[62] P. Sagetong and A. Ortega. Channel adaptive multiple description coding for image transmission over packet loss channels. *In preparation for IEEE Trans.on Image Communication.*

[63] P. Sagetong and A. Ortega. Optimal bit allocation for channel-adaptive multiple description coding. In *Proceedings of VCIP'2000*, pages 53–63, San Jose, CA, Jan. 2000.

[64] P. Sagetong and A. Ortega. Analytical model-based bit allocation for wavelet coding with applications to multiple description coding and region of interest coding. *IEEE International Conference on Multimedia and Expo (ICME)*, Aug 2001.

[65] P. Sagetong and A. Ortega. Rate-distortion model and analytical bit allocation for wavelet-based region of interest coding. *IEEE 2002 International Conference on Image Processing,ICIP 2002*, Sept 2002.

[66] P. Sagetong and A. Ortega. Message-passing algorithm for two-dimensional dependent bit allocation. *Electronic Image 2003, EI 2003*, Jan. 2003.

[67] P. Sagetong and W. Zhou. Dynamic wavelet-feature-based watermark apparatus and method for digital movies in digital cinema. *HRL Invention Disclosure*, Dec 2001.

[68] P. Sagetong and W. Zhou. Dynamic wavelet feature-based watermarking for copyright tracking in digital movie ditribution systems. *IEEE 2002 International Conference on Image Processing (ICIP), Rochester, New York*, Sept. 2002.

[69] A. Said and W. A. Pearlman. A new fast and efficient image codec based on set partitioning in hierarchical trees. *IEEE Trans. on Circuits and Systems for Video Technology, vol. 6, no. 4, pp. 243-250*, June 1996.

[70] S. D. Servetto, K. Ramchandran, V. Vaishampayan, and K. Nahrstedt. Multiple-description wavelet based image coding. In *Proceeding of ICIP-98*, volume 1, pages 659–663, Chicago, IL, Oct 1998.

[71] J.M. Shapiro. Embeded image coding using zerotrees of wavelet coefficients. *IEEE Trans. Signal Processing, vol. 41, pp. 3445-3462*, Dec 1996.

[72] P.G. Sherwood, X. Tian, and K. Zeger. Efficient image and channel coding for wireless packet networks. *Proc. of Intl. Conf. on Image Proc., ICIP'00*, 2000.

[73] Y. Shoham and A. Gersho. Efficient bit allocation for an arbitrary set of quantizers. *IEEE Trans. Acoust. Speech Signal Process., ASSP-36(9), pp. 1445-1453*, Jan. 1988.

[74] R. Singh and A. Ortega. *Consistency estimation of erased data in a DPCM based multiple description coding system.* Kluwer Academic Publishers, 2002.

[75] R. Singh, A. Ortega, L. Perret, and W. Jiang. Comparison of multiple description coding and layered coding based on network simulations. In *Proceedings of VCIP'2000*, San Jose, CA, Jan. 2000.

[76] D.R. Stinson. *Cryptography, Theory and Practice.* New York: CRC Press, 1995.

[77] M. D. Swanson, B. Zhu, and A. H. Tewfik. Multiresolution scene-based video watermarking using perceptual models. *IEEE Journal on selected areas in communications, vol. 16, no. 4, pp. 540-550*, May 1998.

[78] D. S. Taubman and M. W. Marcellin. *JPEG2000: Image compression fundamentals, standards and practice.* KAP, 2002.

[79] P. Thiennviboon, K. Chugg, and A. Ortega. Simplified grid message-passing algorithm with application to digital image halftoning. *IEEE 2001 International Conference on Image Processing (ICIP 2001), Thessaloniki, Greece*, Oct. 2001.

[80] P. Thiennviboon, K.M. Chugg, and A. Ortega. Model-based digital image halftoning using iterative reduced-complexity grid message-passing algorithm. *SPIE Proceedings Image and Video Communications and Processing 2003, EI2003*, Jan. 2003.

[81] C.-K. Toh, M. Delwar, and D. Allen. Evaluating the communication performance of an ad hoc wireless network. *IEEE trans. on wireless communications, vol. 1, issue 3, pp. 402-414*, July 2002.

[82] Y. Wang, M. Orchard, and A. R. Reibman. Multiple description image coding for noisy channels by paring transform coefficients. *Proceeding of MMSP-97*, June 1997.

[83] Y. Wang, J. Ostermann, and Y. Zhang. *Video processing and communications.* Prentice hall, 2001.

[84] Z. Wang and A. C. Bovik. Bitpland-by-Bitplane Shift (BbBShift)- a suggestion for JPEG 2000 region of interest image coding. *IEEE Signal processing letters, Vol. 9, No. 5, pp. 160-162*, May 2002.

[85] J. Watkinson. *The MPEG handbook : MPEG-1, MPEG-2, MPEG-4.* Focal Press, 2001.

[86] J. Wei, Y. Hagihara, and H. Kobatake. Detection of cancerous tumors on chest x-ray images - candidate detection filter and its evaluation. *ICIP 1999 Proceeding, vol. 3, pp. 397-401*, 1999.

[87] R. B. Wolfgang, C. I. Podilchuk, and E. J. Delp. Perceptual watermarks for digital images and video. *Proceedings of the IEEE, Speical Issue on Identification and Protection of Multimedia Information, vol. 87, no. 7, pp. 1108-1126*, July 1999.

[88] S.-W. Wu and A. Gersho. Rate-constrained optimal block-adaptive coding for digital tape recording of hdtv. *IEEE Trans. on Circuits and Sys. for Video Tech., 1(1):100-112*, Mar. 1961.

[89] W. Zhou, T. Rockwood, and P. Sagetong. Non-repudiation oblivious watermarking schema for secure digital cinema distribution. *IEEE 2002 International Workshop on Multimedia Signal Processing, MMSP 2002*, 2002.

[90] W. Zhou, A. Vellaikal, Y. Shen, and C.-C J. Kuo. On-line scene change detection of multicast video. *IEEE Journal of visual communication and image representation, vol. 16, no. 4, pp. 540-550*, Jan 2001.