Techniques for Image Enhancement and Robust Video Communications

by

CHAN Chi-Lun

A dissertation submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

in the Department of Electronic and Information Engineering

The Hong Kong Polytechnic University

Hong Kong SAR, China

February 2003

Abstract
of thesis entitled
"Techniques for Image Enhancement and Robust Video Communications"
submitted by Chan Chi-Lun for the degree of Doctor of Philosophy
at The Hong Kong Polytechnic University in February, 2003.

With the improvement in communication technology, broadband services are readily available for general multimedia communications. While the bottleneck in communication bandwidth has been relieved, the quality of the delivered images and videos becomes the next major concern when designing the communication systems. In this research work, we consider two important topics as far as quality image and video communications is concerned: 1) image enhancement; 2) source coding schemes for reliable transmission.

Current studies on image and video communications consider only the problem due to the transmission channel. To achieve quality image and video delivery, the problems incurred when acquiring the images or videos should not be ignored. In this work, we investigated two major kinds of image or video acquisition problems: blurring and noise contamination. We study the enhancement techniques for such problems. In particular, to deal with the blurring problem of acquired images, a fast algorithm is proposed for blind image restoration based on the discrete periodic Radon transform (DPRT). With the DPRT, the original 2-dimensional blind image restoration problem is converted into 1-dimensional ones such that the complexity is much reduced and thus faster restoration can be achieved. Besides, to enhance the performance of compression with noisy acquired image, we investigate the adaptive thresholding technique that is particularly suitable for the MPEG-4 still texture image codec in coding noisy images. It integrates the denoising process and the coding process into a single step for the traditional MPEG-4 still image coder such that the visual quality of the decoded images are much improved.

For video communication, a novel model is proposed to describe the relationship between the quantization and the rate-distortion (RD) performance of block-based video

codec. We show that the distortion-quantization relationship, $D(Q)$, of most coded sequences, encoded with different group-of-picture structures, can be approximated by a single linear model using regression analysis. We further verify the rate-quantization relationship, $R(Q)$, based on the proposed $D(Q)$ model. Estimates from the proposed models are compared with actual measurements and promising results are obtained. Furthermore, to realize robust transmission, a new error resilient transcoding scheme for H.263 video transmission over combined wireless and wired channel is proposed. It first determines the coding type of every group-of-block in each frame during the coding process by considering the RD performance of the coder. A transcoder is then proposed to assay the transmitted bitsteam and revise the corrupted parts in case of error. Finally, the revised bitstream is sent to the decoder through a reliable channel. As compared with the traditional error resilient approaches, the proposed scheme greatly improves the quality of the decoded video.

Acknowledgements

I would like to take this opportunity to express my sincere gratitude to various bodies. I am eternally grateful to my supervisor, Dr. Daniel Pak Kong Lun, for his patience guidance, continuous support, and constant encouragement throughout my study. It is a rare privilege and a great honour for me to work with him.

I would like to thank my friends and colleagues for enriching my life. In particular, my thanks are due to Dr. T. C. Hsung, Dr. K. P. Wong, Dr. W. F. Poon, Dr. W. C. Poon, Dr. W. K. Fong, Mr. K. H. Ho, Mr. W. K. Cheuk, Mr. K. K. Yiu, Mr. K. Y. Choi, Mr. K. C. Lai, and those who have worked with me during the last four years. The countless discussions with them have made my life and research work an enjoyable experience.

Thanks also to all members of staff of the department of Electronic and Information Engineering, especially, Professor W. C. Siu and the clerical staff in the General Office. They have created a supportive environment for me to work in. It is also my pleasure to acknowledge the Research office of the Hong Kong Polytechnic University for its generous support over the previous years.

Last but not least, I deeply thank my family and Miss. W. L. Chong for their understanding and perseverance.

Notations

$\mathbf{R}$	Set of real numbers.
$\mathbf{R}^2$	$\mathbf{R} \times \mathbf{R}$.
$\mathbf{Z}$	Set of integers.
$\mathbf{Z}^2$	$\mathbf{Z} \times \mathbf{Z}$.
$\mathbf{Z}_n$	$\{0, 1, 2, \ldots, n\text{-}1\} \subset \mathbf{Z}$.
$\lvert f \rvert$ or $abs(x)$	Absolute value of f.
Conj(f)	Complex conjugate of f.
$E[f]$	Expectation of f.
$\lfloor f \rfloor$	Integral part of $f \in \mathbf{R}$.
$\lVert f \rVert^2$	Norm of f.
$L^P(\Omega)$	Class of all measurable functions u, defined on Ω, for which $\int_\Omega \lvert u(x) \rvert^P dx < \infty$.
$l^P(\Omega)$	Class of all measurable functions u, defined on Ω for which $\sum_\Omega \lvert u(x) \rvert^P < \infty$.
$< a >_P$	The positive residue of $a \in \mathbf{Z}$ modulo a positive integer $P \in \mathbf{Z}$.
$f ** g$	2 dimensional linear convolution of the function f with the function g.
$f \otimes g$	1 dimensional circular convolution.
$f \otimes_2 g$	2 dimensional cyclic convolution.

List of Abbreviations

ARMA	Auto-regressive moving average.
B-frame	Bi-directional-coded frame.
BER	Bit-error-rate.
CWT	Continuous wavelet transform.
DCT	Discrete cosine transform.
DPRT	Discrete periodic Radon transform.
DWT	Discrete wavelet transform.
FFT	Discrete fast Fourier transform.
GCV	Generalize cross-validation.
GOB	Group-of-block.
GOP	Group-of-picture.
I-frame	Intra-coded frame.
JPEG	Joint picture experts group.
MB	Macroblock.
MBRM	Macroblock resynchronization marker.
MPEG	Moving picture experts group.
MSE	Mean-square-error.
N-D	N-dimensional.
P-frame	Predictive-coded frame.
PDF	Probability density function.
PSNR	Peak signal-to-noise ratio.
PSF	Point spread function.
RD	Rate-distortion.
RM	Resynchronization marker.

Statement of Originality

The following contributions reported in this dissertation are claimed to be original.

1. The efficient algorithm for blind image restoration via the discrete periodic Radon transform (DPRT) (Chapter 3, Section 3.4.2).

   This is a new algorithm to improve the performance of using 2-D generalize cross-validation (GCV) in blind image restoration. By this algorithm, as much as 96% saving in arithmetic operations can be achieved over the conventional 2-D GCV approach in blind image restoration. It is achieved by transforming the observed image into the DPRT domain, then converting the 2-D GCV algorithm into several 1-D ones in restoring the point spread function (PSF) and the original image.

2. Analysis on the effect of noise to the MPEG-4 still texture image coding algorithm (Chapter 4, Section 4.2).

   We investigate the effect of Gaussian noise to the MPEG-4 still texture image compression scheme. We show that the quantization process can reduce the effect of noise to a certain extent. Besides, we explain why there is an optimal coding point when using the codec to compress a noisy image.

3. The wavelet-based adaptive thresholding technique for improving the performance of the MPEG-4 still texture image codec in coding noisy images (Chapter 4, Section 4.4).

   This is a new approach to carry out thresholding technique in image denoising by considering the effect of quantization during compression process. By this technique,

the visual quality and the peak signal-to-noise ratio (PSNR) of the reconstructed images are much improved comparing with the traditional MPEG-4 approach in coding noisy images. It is achieved by utilizing and comparing the probability density function of the detailed level wavelet coefficients and noise. Threshold values are selected according to the quantization step sizes for reducing noise during the coding process.

4. The novel group-of-picture (GOP) based rate-distortion (RD) modeling for block-based video compression (Chapter 4, Section 5.4).

   This is a new scheme to determine the RD relationship for a sequence of block-based coded video frames. By this modeling scheme, the relationship between the quantization and the RD performance of block-based compressed videos can be estimated accurately. The proposed models are derived based on the regression analysis and the parameters of the models are estimated using a least-square optimization approach.

5. The new macroblock (MB) level resynchronization marker (MBRM) for robust H.263 communications (Chapter 6, Section 6.3.1).

   This is a new resynchronization marker (RM) introduced for error resilience. This new RM contains a resynchronization word (12 bits) and an MB number indicator (4 or 5 bits depends on QCIF or CIF format) for bitstream resynchronization in case of error. As compared to other traditional RMs which operate at the group-of-block or slice levels, the proposed MBRM works in the MB level and it localizes error in an effective way so that accurate concealment and resynchronization can be performed in case of error.

6. The group-of-block (GOB) based coding-mode-selection algorithm for error resilient H.263 coding (Chapter 6, Section 6.3.2).

This is a novel source coding algorithm for error resilient H.263 coding. By this algorithm, the average loss (measured in terms of PSNR) in the reconstructed frames can be kept within 2 dB while the other approaches give 5-8 dB loss when transmitting compressed videos using H.263 over an error-prone channel with bit-error-ratio (BER) ranged from 10^{-3}-10^{-5}. It is achieved by jointly optimizing the use of MBRM and intra-refreshing technique in determining the coding mode of each GOB in a video frame.

Table of Contents

List of Figures

List of Tables

Author's Publications

International Journal Papers

1. Daniel P. K. Lun, Tommy C. L. Chan, T. C. Hsung, David D. Feng, and Y. H. Chan, "Efficient Blind Image Restoration using Discrete Periodic Radon Transform," paper accepted by *IEEE Trans. on Image Processing*, USA.

2. Tommy C. L. Chan and Daniel P. K. Lun, "Improved MPEG-4 Still Texture Image Coding under Noisy Environment," paper accepted by *IEEE Trans. on Image Processing*, USA.

3. Tommy C. L. Chan, and Daniel P. K. Lun, "Optimally Using Mode-selection and Resynchronization Marker for Error Resilient Video Coding," paper submitted to *IEE Proceedings, Vision, Image and Signal Processing*, UK.

International Conference Papers

4. T. C. Hsung, Tommy C. L. Chan, Daniel P. K. Lun and David D. Feng, "Embedded Singularity Detection Zerotree Wavelet Coding," *Proceedings, IEEE International Conference on Image Processing* (ICIP'99), vol. 2, pp. 274-8, Kobe, Japan, October 1999, Japan.

5. Tommy C. L. Chan, Daniel P. K. Lun and Z. N. Zhang, "Lossy Compression for Noisy MPEG-4 Still Texture Image," *Proceedings, International Symposium on Signal Processing and Intelligent System* (ISSPIS'99), pp. 648-651, Guangzhou, China, 1999.

6. Tommy C. L. Chan and Daniel Pak-Kong Lun, "Adaptive Thresholding for Noisy MPEG-4 Still Texture Image", *Proceedings, IEEE International Conference on Image Processing* (ICIP'2000), vol. 3, pp. 833-836, BC, Canada, 2000.

7. Daniel P. K. Lun, Tommy C. L. Chan, and David D. Feng, "Improved Non-Invasive Quantification of Physiological Processes with Dynamic PET using Blind System Identification," *Proceedings, International Symposium on Intelligent Multimedia, Video and Speech Processing* (ISIMP'2001), pp. 255-258, Hong Kong, May 2001.

8. Daniel P.K. Lun, Tommy C.L. Chan, T.C. Hsung, and David D. Feng, "Efficient Blind Image Restoration Based On 1-D Generalized Cross Validation," *Proceedings, Advances in Multimedia Information Processing* (PCM'2001), Springer-Verlag, New York, pp. 434-441, 2001.

9. Daniel P. K. Lun and Tommy C. L. Chan, "Optimal Mode-Selection Scheme for Error Resilient Video Coding," paper submitted to *IEEE International Conference on Image Processing* (ICIP'2003), Barcelona, Spain.

Chapter 1 Introduction

1.1 Overview

In the past decade or so, the spectacular advances in communication and computer technology have resulted in a rapid growth in data communications, especially the multimedia communications. Multimedia can be generally referred as the presentation of information through multiple forms of media including audio, animation, graphic, image, speech, text, video data, and etc. Multimedia communications provide the possibility for a spectrum of new applications, such as videotelephony, videoconferencing, remote surveillance, and remote shopping, in the areas of information distribution, and cooperative work and have been definitely changing the everyday life of people.

Due to the relatively high demand in storage and computing resources, it is not surprised to see that, among the mixed media information involved in a multimedia communication system, the efficient transmission of image and video signals have attracted much attention particularly. Typical architecture for most digital image/video communication systems can conventionally be defined as the assembly of the modules in Fig. 1.1.

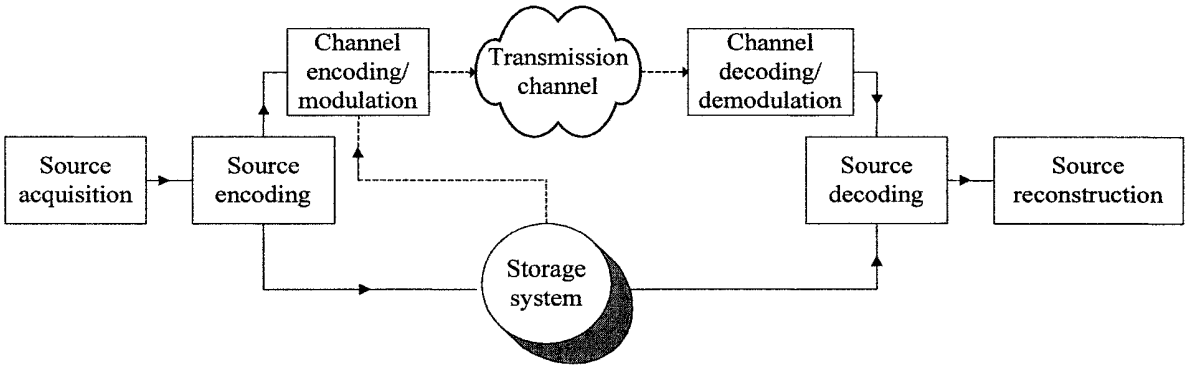


Figure 1.1: Typical system architecture for most communication systems.

Basically, a video frame can be regarded as an image and most images originate in an optical form [1]. As the very first step of a communication system as shown in Fig. 1.1, the source acquisition process converts an optical source, no matter it is an image or video frame, to an electrical signal with some sensor devices. Different types of sensors exist. For example, the charge coupled device (CCD) video cameras [1] is one of the commonly used sensors that generates electrical signal to represent the image acquired. For digital communication systems, the analog electrical signal needs to be digitized and turned into a digital data form using an analog-to-digital (A/D) converter [2] before they can be stored or processed by a digital computer. After the digitization process, an image (or video frame) is represented by discrete points of numerically defined brightness, called pixels or voxels. By manipulating these brightness values, different complex operations, for example, image enhancement, restoration, analysis, and compression, can easily be carried out by using a digital computer. Particularly, as the performance of most communication systems depend very much on the quality of their source image, image enhancement becomes an important step to improve the overall performance of most communication systems. Actually, image

processing operations can also be applied to an image source in the optical or analog form. However, these issues will not be discussed in this dissertation.

Most acquired images, especially video frame sequences, inherently contain large amount of redundant information. In order to improve the transmission efficiency of image or video data over the band-limited channel or the efficiency of their storage capacity, source compression, which tries to eliminate the redundancies within the image or video source, becomes the most compelling for most communication systems. There are two general forms of compression: lossless and lossy compression. Lossless compression techniques preserve the exact data found in the original source. Run-length coding [3] and Huffman coding [4] are two examples of lossless compression. Lossy compression techniques do not exactly represent the data of the original source but strive only to maintain a particular level of subjective quality. Typical examples of the lossy compression technique are uniform quantization [5] and vector quantization (VQ) [6]. By combining the lossless and lossy compression techniques, several well-known hybrid encoding schemes, such as JPEG [7], JPEG-2000 [8], and embedded zero-tree wavelet (EZW) image coding [9, 10], H.263 [11], H.263+ [12], and MPEG-4 [13] video coding schemes, have been developed. Compressed image/video is then either stored using a storage device or fed to a channel encoder stage for further processing before it is transmitted. Depending on the application feature, some applications, for example the Video-On-Demand (VOD) [14], require to store the compressed bitstream for later distribution rather than transmitting it directly to the receiver/decoder side. The storage media can be either in magnetic or optical form. Magnetic data storage carriers [15] take the form of floppy disks or hard-disks. They provide low average access time and adequate capacity for a reasonable price. However, due to the higher storage capacity and lower cost, optical storage media for example the Compact Disk-Read Only Memory (CD-ROM) [16], are much preferred for most applications.

For the channel coding stage, the compressed bitstream is segmented into fixed or variable-length packets followed by an error protection process before transmission. Typical error protection processes use forward error correction (FEC) codes [17, 18] and interleaving [19] to protect the bitstream from transmission errors. If the transmission channel being adopted is a wireless or mobile one, digital modulation [20] is further applied to map digital data from the bitstream to the corresponding radio frequencies suitable for transmission over the channel. At the receiver end, the reverse operations are performed accordingly to reconstruct the original images or videos. Different from the one-way communication system as shown in Fig. 1.1, some video communication systems, for example video conferencing systems, may allow data flowing in the other direction to convey some feedback information.

In fact, error control in the processing and the transmission of good quality image and video data through unstable and band-limited channel is extremely crucial to many typical wireless or mobile multimedia communication systems. In this dissertation, we focus on the aspects related to image enhancement and source coding schemes for quality video communications.

1.2 Problem Addressed

The quality of the received images and videos of most practical communication systems, including the one shown in Fig. 1.1, is determined by two criteria: the quality of source and the process of transmission.

An image or video acquisition system usually includes a lighting system, lens, photodetector, and a digitization unit. Each of these components can contribute degradation to the final images. Therefore, the process of source acquisition is never error-free and we

cannot expect to get a perfect image out of a natural scene. In general, acquired image or video frame consists of a degraded representation of the original object and the degradation is due to either the process of image formation or the process of image recording [21].

The degradation due to the process of image formation is usually denoted by blurring. Motion blurs and out-of-focus blurs are two examples of blurring in most cases. Motion blurs [22] are caused by relative motion, during exposure, between the camera and the object being photographed. Out-of-focus blurs are due to the lack-of-focus of the camera lens [23]. Blurring is a deterministic process and can be well described using mathematical models. On the other hand, the degradation introduced by the recording process is usually denoted by noise. Noise is generated usually due to the measurement errors of capturing devices and the random disturbances introduced by the acquisition environment. Common noise sources contained in the camera itself are photon noise, thermal noise, and readout noise [24]. Noise is a statistical process so that the noise affecting a particular image is unknown but it can still be modeled by adopting certain assumptions. The characteristics of noise depend on their source and they can generally be grouped into two classes: independent noise and noise which depends on the image data. In reality, degradation introduced by source acquisition not only degrades the source quality to a great extent but also degrades the performance of most coding scheme [25] due to the extra bits needed to code the noise.

The source of degradation in the transmission process of a communication system is usually owing to the channel errors. Channel error can be further classified into two categories [26]: random bit errors and erasure errors. Random bit errors are caused by the imperfection of physical channels, which results in bit inversion, insertion, and deletion. Erasure errors can be caused by packet loss in packet network such as the Internet, burst errors in the storage media due to physical defects, or system failures for a short period of time.

In a digital communication system, the visual content is usually divided into smaller units, such as blocks of pixels, then compressed and encoded, and finally transmitted over a fixed or variable capacity channel. Channel errors can severely degrade the performance of the system. Since most digital video compression methods are not memoryless, the impact of channel errors is more severe than in the analog case and may extend spatially and temporally. Based on the coding methods adopted in the source encoder and the affected visual content, the impact of random bit errors can range from negligible to objectionable. When fixed-length coding is used, a random bit error affects only one codeword and the damage caused is generally acceptable. However, due to the extensive use of variable length coding (VLC), such as the Huffman coding, in the modern image and video codecs, random bit errors can desynchronize a coded bitstream in such a way that many following bits are undecodable until the next resynchronization point appears. This loss of synchronization between encoders and decoders caused by the random bit errors in VLC coded bitstreams can also lead to effective erasure errors. The effect of erasure errors is much more destructive than random bit errors, due to the loss of or damage to a contiguous segment of bits. However, both the random bit errors and erasure errors can degrade the reconstruction quality quickly. Unless the encoders and decoders take proper remedial steps, the communication system can totally break down.

1.3 Related Work

Many approaches have been proposed to solve the problems of blurred and noisy image and the erroneous transmission of image or video bitstreams. Basically, previous methods to solve the above problems can be classified into three categories, namely: 1) blurred image restoration; 2) image noise reduction and 3) error resilient techniques for image/video

communications. In this section, we outline some of the related works in each category synoptically.

1.3.1 Blurred image restoration

Due to the degradations caused by the imperfection of the image formation and recording processes for most communication systems, it is very often unavoidable to get a degraded image or video frame during source acquisition. What we can do is to recover an original source or properly speaking, the approximation of the original source, from the degraded observations. It can be done by modeling the degradations and applying an inverse procedure to obtain an approximation of the original scene. Classical techniques in this area, such as inverse filtering [27], Wiener filtering [28], and constrained least-squares (LS) filtering [29], can give good performance by assuming that the information of the blur or noise is known a priori. However, in practice, the information of blur or noise is unknown such that the restored sources have to be identified directly from the observations by using partial or very minimal information about the blurring process and the true sources. This restoration process is the so-called blind image restoration [30, 31, 32]. One of the important classes of approach for solving the blind restoration problem is based on the auto-regressive moving average (ARMA) parameter estimation method which tries to model the true image as a two-dimensional (2-D) auto-regressive (AR) process and the blurring as a 2-D moving average (MA) process. Based on these models, the resulted blurred image is represented as an ARMA process which can be modeled using the ARMA input-output equations with space varying parameters [31, 33]. Identifying the ARMA parameters allows us to identify both the true image and the blur.

The existing methods of this class differ in how the ARMA parameters are estimated. Techniques based on second-order statistics, such as maximum-likelihood (ML) estimation [34, 35, 36], generalized cross-validation (GCV) [37, 38], and neural networks

[39] have been proposed. Higher order statistics methods have also been used for ARMA parameter estimation [40]. Among them, the ML estimation and GCV approaches are the most successful for blurred image restoration.

The ML approach has a long history for ARMA parameter estimation. It determines the various parameters that produce the blurred image on hand. The use of expectation-maximization method [41] in the ML approach allows the parameter estimation to become straightforward and efficient. The GCV approach on the other hand determines the parameters that minimize a weighted sum of prediction errors. This criterion has been shown to possess certain properties that are superior to those of ML in the context of regularization parameter estimation [42]. Nevertheless, the GCV approach is often suffered from heavy computational load due to the extensive numerical operations required to search for the minimum GCV score.

1.3.2 Image noise reduction

Apart from blurring, noises introduced by the camera itself and the capturing environment (background noise) can also be the main source of degradation for the acquired images or video frames. Actually, image noise reduction has been studied for many years [28]. Basically, classical approaches for image noise reduction tend to use filtering techniques due to its simplicity.

Classical filters for noise reduction can be further classified into two types: linear filters and nonlinear filters. A linear filter replaces the gray-value of a pixel by a linear combination of its neighborhood and it is efficient to reduce additive Gaussian noise. A mean filter, for example, is the optimal filter for Gaussian noise in the sense of minimum mean square error. Linear filters, however, tend to blur sharp edges, destroy lines and other fine image details, fail to effectively remove heavy-tailed noises, and perform poorly in the presence of signal-dependent noise. This led to a search for nonlinear filtering as an

alternative. The research effort on nonlinear median-based filtering [43], for example, has achieved remarkable results and has highlighted some new promising research avenues. The median filter, however, often tends to remove fine details in the image, such as thin lines and corners. In recent years, a variety of median-type filters such as stack filters and weighted median filters [44] have been developed to overcome this drawback. In fact, filtering techniques can be applied in both the spatial domain and the frequency domain for noise reduction.

Frequency filters, such as low-pass filters [45], process an image in the frequency domain. The image is Fourier transformed [1], multiplied with the filter function and then inverse-transformed into the spatial domain. Attenuating high frequencies results in a smoother image in the spatial domain, attenuating low frequencies enhances the edges. All frequency filters can also be implemented in the spatial domain if there exists a simple kernel for the desired filter effect. However, it is computationally less expensive to perform the filtering in the spatial domain. Frequency filtering is more appropriate if no straightforward kernel can be found in the spatial domain.

Each filtering technique makes its own assumptions depending on the type of imagery and the goals of the restoration. Hence, one technique is not useful for all applications. Many other techniques based on mathematical morphology [46], fuzzy control rules [47], partial differential equations [48], and wavelet theory [49, 50] have been developed for noise reduction purpose as well.

In particular, due to the idea of multiresolution and the ability to localize time-frequency information, noise reduction using wavelet methods [51, 52, 53, 54] has aroused much attention recently. Many of these algorithms use the properties of wavelet decomposition to distinguish the regular part of a signal from noise and have been proven to work well in many applications.

1.3.3 Error resilient techniques for image and video communications

Many resilience mechanisms have been proposed in the literature to protect digital visual communications against channel errors and they can be categorized into four classes:

- Methods carried out by mainly the source encoder.
- Joint source and channel coding
- Post-processing executed at the decoder to conceal errors
- Interaction between encoder and decoder

For the first class of methods, the source encoder plays the primary role to add intentional redundancy to the bitstream. The redundant data are carefully designed to achieve maximum gain in error resilience with minimum performance penalty in coding efficiency. Well-known examples of such techniques include insertion of resynchronization markers (RM) [55], reversible variable-length coding (RVLC) [56, 57], insertion of intra-coded blocks or frames [58], data partitioning [55, 59], error-resilient entropy code (EREC) [60, 61], and forward error correction coding [62].

Joint source and channel coding is accomplished by co-designing the source encoder and the channel encoder based on some given channel error characteristics in order to minimize the effect of transmission errors. Most techniques in this class try to solve the problem of bit allocation between the source encoding and channel encoding such that the overall reconstruction distortion in the receiver end is minimized [63, 64]. Traditional multiple description coding (MDC) [65] and layered coding (LC) with unequal error protection [66] are considered to belong to this class.

The third class of methods can be referred as error concealment methods that are implemented in decoders only without any interaction between the encoder and the decoder. A variety of techniques have been proposed for the recovery of texture information. Most of them perform either simple interpolation [67] or motion-compensated temporal prediction [68]. Some of them try to recover the corrupted region using prior information about the

underlying image structure [69]. Usually, the recovery of missing motion vectors is based on simple heuristics or mean/median filtering [70].

Methods in the fourth class require a feedback channel from decoder to encoder. If feedback channel exists, information about the corruption of bitstream due to errors can be sent from decoder to encoder by means of automatic repeat request (ARQ) such that encoder can adjust its operation, for example, to retransmit the lost data, to suppress the effects of the errors. However, this will incur processing delays and may be unacceptable for certain real-time interactive applications [71]. Another way to solve the problem is to permit errors but limit their propagation. Reference picture selection (RPS) [12] is a method in this class that when information about errors in a previously encoded frame is sent from decoder to encoder, encoder selects an undamaged frame as reference for the prediction. However, this approach requires larger buffer size in both the encoder and the decoder. Some sophisticated approaches track the error effect across the reconstructed frames in decoder, by means of error tracking [72], such that encoder can code the corresponding damaged blocks in intra-mode or to avoid performing prediction based on the damaged areas. A more extensive review of error resilience techniques can be found in [71, 73].

1.4 Objective

The next generation image and video communications will mainly focus on the reconstructed picture quality. The original source quality and the control of error for the transmission of compressed source through error-prone channels are the two most important factors to determine the reconstructed picture quality. The objective of this research work is to develop new techniques for image enhancement and robust video communication. To achieve the objective, novel techniques for efficient blurred image restoration, image noise

reduction for wavelet-based image coding, and video source coding schemes for reliable transmission were studied and developed.

1.4.1 Blurred image restoration

Traditional image restoration techniques achieve good performance by assuming that the information of the blur/noise or original image is known a priori. Nevertheless, this assumption is practically impossible such that the restored sources have to be identified directly from the observations by using partial or very minimal information about the blurring process and the original sources.

By assuming that the blurring process is a convolution operation between the blur function and the real source, image restoration can indeed be considered as a blind deconvolution operation. However, this is a well-known ill-posed problem in image processing. Although many approaches have been proposed, various problems are incurred in actual implementation due to the large data size and long computational time of these approaches. They are undesirable even with the current computing machines. Hence, an efficient algorithm is needed to reduce the memory size and computational time required for restoring the original image but at the same time retaining the quality of the restored image.

1.4.2 Image noise reduction for wavelet-based image coding

Source acquisitions for many practical image processing systems are done in an outdoor environment that very often background noise is introduced to the acquired images. Together with the noise introduced by the capturing devices, it is rather difficult, if not impossible, to ensure the acquired images to be noise-free. When coding that kind of noisy images, the performance of most coding schemes can be greatly degraded and the reconstructed images at the receiver side will also be corrupted. The recently proposed wavelet codecs [9, 10, 13] are no exception. When coding the noisy images by using the

wavelet-based codec, noise invades the encoder by introducing more large amplitude wavelet coefficients in detailed levels. Hence, extra bits have to be used to encode the noisy coefficient and results in poorer compression efficiency.

It is obvious that if we can distinguish and remove the noisy wavelet coefficients during compression, the problem as mentioned above will be resolved. However, existing noise reduction techniques try to reduce noise without considering its effect to compression. Residual noise that leaves behind after the noise reduction process would be amplified by the compression process and causes serious degradation to the compression efficiency as well as the visual quality of the reconstructed images. Hence, it is necessary to develop a simple but efficient technique for noise reduction by considering the effect of noise on the compression process such that the degradation caused by the residual noise can be minimized.

1.4.3 Video source coding schemes for reliable transmission

It is obvious that if feedback channel is available, error can be recovered accurately by adopting ARQ or precise error tracking techniques. However, providing feedback channel can be expensive in certain applications. Besides, it can also be affected by the round-trip delay of the network and channel errors. Therefore, error resilient coding technique without using feedback information becomes an increasingly important research topic. Many error resilient tools have been proposed for reliable video communications without feedback. For different level of error protection requirements, some approaches aim at isolating errors, some of them are proposed to facilitate error detection and error concealment in decoder, while the others try to recover the corrupted sources. A good choice of error resilient tools, or a combination of them, is very crucial not only to the protection of the compressed sources but also the efficiency of the communication systems.

One of the most important problems under investigation is how to effectively make use of the existing techniques to achieve maximum protection and coding efficiency to the bitstream. To this problem, it is essential to develop a robust coding scheme by carefully considering the feature of the error resilient tools, effect of error propagation and also the error concealment method being adopted in decoder. This research work tries to look into this problem and provide the solution.

1.5 Dissertation Outline

This dissertation is organized as follows:

In Chapter 2, a general review on the subject areas which are relevant to our present work is given. They include the techniques for restoring blurred image by using ARMA parameter estimation and GCV, image denoising based on wavelet-based methods, and the related techniques for error resilient image/video communications.

In Chapter 3, we present a concise review on the properties and reconstruction method of the discrete periodic Radon transform (DPRT) [74, 75]. In addition, the application of DPRT to efficient blind image restoration is discussed.

In Chapter 4, the effect of noise to the wavelet-based MPEG-4 still texture image codec is first elaborated. A novel adaptive thresholding technique is then proposed to enhance the performance of the corresponding codec in coding noisy acquired images.

In Chapter 5, we verify and propose a novel model to describe the relationship between the quantization and the RD performance of block-based video codec.

In Chapter 6, a novel video source coding scheme is proposed by jointly optimizing the uses of coding-mode-selection and resynchronization marker (RM) for error resilience. Potential applications of the proposed scheme are also discussed.

Finally, the work presented in this dissertation is summarized in Chapter 7. Future extensions of the present work are also discussed.

Chapter 2 General Review on Relevant Techniques

2.1 Introduction

In this chapter, we review some existing approaches which are relevant to the present work. Firstly, the background of image restoration and a summary of blind restoration of blurred image using the techniques of auto-regressive moving average (ARMA) modeling and generalized cross-validation (GCV) parameter estimation are given. Subsequently, theory of wavelet transform and its applications to image noise reduction are discussed. Finally, a brief overview for error resilient image/video communications is provided. In particular, the following topics are discussed. They include the techniques for current image/video transform coding; types of channel errors and their effects to compressed image/video bitstream; techniques for error detection; and the current error resilient tools for source coding.

2.2 Restoration of Blurred Image by Auto-Regressive Moving Average (ARMA)

In digital image processing, the degradation of true image due to blurring and additive noise during the acquisition process can be modeled by a linear degradation model as illustrated in Fig. 2.1:

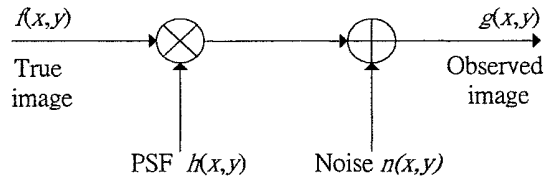


Figure 2.1: Linear degradation model.

With the notations in Fig. 2.1, the observed image can be formulated as:

$$g(x,y) = \sum_{(k,l)\in R_h} h(k,l) f(x-k, y-l) + n(x,y) \tag{2.1}$$

where $f(x,y)$ and $g(x,y)$ denote the original unblurred image and the observed image, respectively. $h(x,y)$ presents a 2-D linear shift-invariant blur function, which is also known as the point-spread function (PSF), with a finite support R_h and $n(x,y)$ is the additive noise introduced by the imaging system. Without loss of generality, $f(x,y)$ is usually assumed to be positive (i.e., real and non-negative) and has a finite support. It can be shown from Eq. (2.1) that the acquired image can be formulated in terms of the 2-D linear convolution of $f(x,y)$ and $h(x,y)$ as:

$$g(x,y) = h(x,y) ** f(x,y) + n(x,y) \tag{2.2}$$

where ** represents the 2-D linear convolution operator. The problem of restoring the unknown original image $f(x,y)$ requires the deconvolution of the PSF $h(x,y)$ from $g(x,y)$. If $h(x,y)$ is assumed to be known explicitly prior to the deconvolution procedure, many well-known classical linear image restoration techniques, as mentioned in Section 1.3.1, can be applied to predict the original image $f(x,y)$. Nevertheless, in many practical situations, the information of blur is usually undetermined. In this case, we have to estimate $f(x,y)$ directly from $g(x,y)$ by using partial or very minimal information about $h(x,y)$ or $f(x,y)$. Such an estimation problem, with the linear degradation model in Fig. 2.1 and Eq.(2.2), is called the blind image restoration. The problem of blind image restoration can be tackled as a single

problem or divided into two steps, that is, identifying the PSF separately and then performing the image restoration. Thorough discussion on the important characteristics and constraints of the problem of blind image restoration can be found in [31, 32].

In Eqs. (2.1) and (2.2), the noise term $n(x,y)$ is usually assumed to be a zero mean Gaussian distributed white noise. Furthermore, the PSFs can be represented by different parametric models based on the nature of degradations. As mentioned in Section 1.2, motion blur and out-of-focus blurs are two typical degradations encountered for most communication systems and their corresponding PSFs can be represented by the following analytical models:

1. Motion blur: represents the 1-D uniform local averaging of neighboring pixels, a common result of camera panning or fast object motion and horizontal motion,

$$h(x) = \begin{cases} \frac{1}{L}, & \text{if} \quad -\frac{L}{2} \le x \le \frac{L}{2} \\ 0, & \text{otherwise.} \end{cases} \tag{2.3}$$

2. Uniform out-of-focus blur: defocusing found in a variety of imaging systems as uniform intensity distribution within a circular disk,

$$h(x,y) = \begin{cases} \frac{1}{\pi R^2}, & \text{if} \quad \sqrt{x^2 + y^2} \le R \\ 0, & \text{otherwise.} \end{cases} \tag{2.4}$$

3. Uniform 2-D blur: A more severe form of degradation that approximated an out-of-focus blur and is commonly adopted in many research simulations,

$$h(x,y) = \begin{cases} \frac{1}{L^2}, & \text{if} \quad -\frac{L}{2} \le i, j \le \frac{L}{2} \\ 0, & \text{otherwise.} \end{cases} \tag{2.5}$$

In Eqs. (2.3), (2.4) and (2.5), L and R are the model parameters to be estimated. Very often, the linear degradation model in Eq. (2.2) is re-ordered lexicographically by stacking either

the rows or the columns of the image into a vector and is rewritten in terms of a matrix-vector form as:

$$g = Hf + n \tag{2.6}$$

where g, f, and n are the observed image, original image and noise, respectively. If the original image is of size $N \times N$, then g, f, and n are represented by vectors with size $N^2 \times 1$, and H will be represented by a $N^2 \times N^2$ matrix [28, 30]. The degradation caused by PSF and $n(x,y)$ is usually measured in terms of a metric called the Blurred Signal-to-Noise Ration (BSNR) as:

$$BSNR = 10 \cdot \log_{10} \left\{ \frac{\frac{1}{MN} \sum_{x,y} \left[G(x,y) - \overline{G}(x,y) \right]^2}{\sigma_n^2} \right\} \tag{2.7}$$

with $G(x,y) = g(x,y) - n(x,y)$,

$\overline{G}(x,y) = E\{G(x,y)\}$

where $E\{x\}$ represents the mean of x, $M \times N$ denotes the image size of $g(x,y)$ and σ_n^2 represents the corresponding additive noise variance.

2.2.1 Blind image restoration using ARMA model

As it is mentioned in Section 1.3.1, one of the important classes of approach to solve the problem of blind image restoration is based on the ARMA parameter estimation method. With such method, the original unblurred image is modeled as a 2-D AR process represented by:

$$f(x,y) = \sum_{\substack{(k,l) \in R_a \\ (k,l) \neq (0,0)}} a(k,l) f(x-k, y-l) + u(x,y) \tag{2.8}$$

where $f(x,y)$ is the original unblurred image pixel value at (x,y), $a(k,l)$ are the AR model coefficients, $u(x,y)$ is the model error which is an independent zero-mean homogeneous

noise process with variance σ_u^2, and R_a is the non-symmetric half-plane (NSIP) support of the AR process [76], respectively. In Eq. (2.8), the parameter $a(0,0)$ equals 1. The AR model coefficients $a(k,l)$ are chosen to minimize σ_u^2. Similar to the conversion between Eqs. (2.2) and (2.6), Eq. (2.8) can be expressed in a matrix-vector form as:

$$f = Af + u. \tag{2.9}$$

The image model in Eq. (2.9) is valid for smooth and homogenous images [77] and it is also suitable for texture images with appropriate model order selection. However, the AR model is not valid for those images that have abrupt changes in local image characteristics, such as edges. A model for blurred image can be obtained by combining Eqs. (2.6) and (2.9) as:

$$g = H(I - A)^{-1}u + n \tag{2.10}$$

where I is the identity matrix. Fig. 2.2 shows the complete blurred image model with respect to Eq. (2.10).

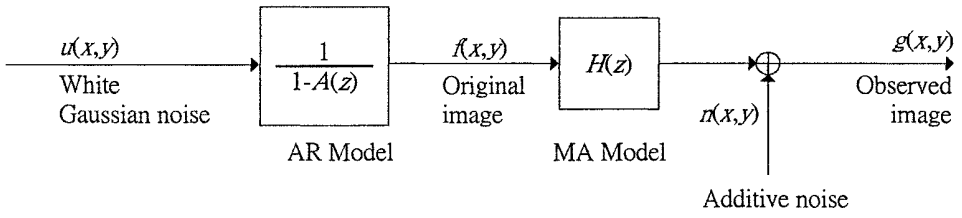


Figure 2.2: A complete blurred image model using ARMA.

In Fig. 2.2, $A(z)$ and $H(z)$ are the Z-transforms of A and H in Eq. (2.10), respectively. By estimating the ARMA model parameters A and H in Eq. (2.10), the PSF and original unblurred image can be estimated. Nevertheless, estimation of A and H in Eq. (2.10) suffers from high computational complexity for PSFs with large support, instability of the estimation algorithms, and non-unique solutions. Therefore, additional constraints are usually made on the PSF during the estimation. For example:

1. The PSF is positive and the mean value of the original unblurred image is preserved in the degradation process such that the number of possible solutions to the problem is limited.
2. The PSF is assumed to be symmetric and zero-phase for the stability and the uniqueness of solution.
3. The PSF has a known parametric form so that the computational complexity of the estimation can be reduced.

Among the many approaches proposed in the literature to estimate the ARMA parameters, the maximum-likelihood (ML) [34, 35, 36] estimation and generalized cross-validation (GCV) [37, 38] are the most successful for image processing applications. Indeed, GCV has higher complexity than ML methods but it is more robust and has been shown to produce better results than the ML approach for real images [38]. In the next section, a brief description of the GCV is given.

2.2.2 Generalized cross-validation (GCV) parameter estimation

General cross-validation (GCV), which is sometimes known as "leave-one-out" or predictive sample reuse, is a well-known technique in the field of data analysis and has been applied as criterion for estimating the optimal regularization parameter in smoothing problem [78].

In GCV estimation, the sample data being tested are divided into two sets: an estimation set and a validation set. The estimation set is used to obtain a model or estimate based on a particular parameter value or assumption. The validation set, on the other hand, is used to validate the performance of the model or estimate and thus the assumption. Therefore, many competing parameter values or assumptions may be tested in order to find the most appropriate set. However, it is quite difficult to divide the data into two sets. On one hand, it is necessary to use as many of the data as possible to obtain a reliable estimate.

On the other hand, it is also desirable to test the estimate on as many of the data that are excluded from the estimation process as possible. GCV tries to solve this problem by arranging and allowing all the data to be used for both purposes. To do so, GCV divides the data into M sets. The model or assumption being tested is imposed on all the sets but one, and a validation error measure is computed from the omitted set. The process continues to select a different set each time until all the M sets have been exhausted for both the estimation and validation process. The validation error measures for each set are averaged to generate the validation error for the particular parameter set or assumption. By this way, all the data are used for both estimation and validation purpose.

Application of GCV for blind restoration of blurred image follows the same procedures as discussed above. Firstly, for a fixed value of model parameters, a temporary restored image is determined using all but one the values from the observed image. The restored image is then reblurred using the corresponding $h(k,l)$ values to predict the observation that was left out of the restoration. During the estimation process, different restorations are performed by leaving out a separate pixel of the observed image each time. The parameter set which provides the minimum mean-square prediction error over all the observations will then be chosen as the optimal image and PSF model parameter estimates.

GCV reduces its sensitivity to additive noise by introducing a regularization parameter to reduce noise amplification in the restoration process. However, the effect of individual parameter to the GCV criterion will become insignificant if the total number of parameters involved in the parameter set is large. Besides, similar to other optimization algorithm, GCV suffers from the problem of local minima. Hence, several assumptions are made on the PSF, as discussed in the previous section, to improve the performance of GCV for image restoration in actual implementation.

2.3 Image Noise Reduction Based on Wavelet Methods

The term "wavelet", which originally called wavelet of constant shape, was first introduced by J. Morlet a decade ago [79, 80]. It is a fairly simple mathematical tool providing successive analysis for many research areas, such as, engineering, mathematics, physics, and many other fields. It provides a common language for the concepts and theories to the algorithms that are suitable, especially, for signal analysis. In order to extract relevant information from a signal for signal analysis, the signal is usually transformed into a different domain using a mathematical transformation. Wavelet transform is one of the transformation tools to change the coordinate system of a signal so that the features hidden in its original domain are revealed in the transform domain. In this section, theory of wavelet, wavelet transform, and its application to image noise reduction are briefly reviewed.

2.3.1 Theory of multiresolution analysis (MRA)

In order to localize time-frequency information in signal analysis, Morlet suggested to analyze a signal with basis functions that are generated by time scaling and translating a function which is well localized in both time and frequency domains. He called the function a "wavelet". The continuous wavelet transform (CWT) is defined as the inner-product of these basis functions and the signal $f(t)$ which is to be analyzed as follows:

$$W_f(a,b) = < f(t), \psi_{ab}(t) > \tag{2.11}$$

where $< x, y >$ denotes the inner-product of x and y with

$$\psi_{ab}(t) = \frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a}\right). \tag{2.12}$$

The function $\psi_{ab}(t)$ is the so-called "mother wavelet" with scaling parameter a and translation parameter b. The dilation and translation properties allow the wavelet basis to analyze signal at different scales and locations. Eq. (2.11) can be rewritten as:

$$W_f(a,b) = \int_{-\infty}^{\infty} f(t)\psi_{ab}^*(t)f(t)dt \tag{2.13}$$

where $y^*(t)$ denotes the complex conjugate of $y(t)$. The continuous wavelet transform performs a mapping of the function $f(t)$ to the scale-space by Eq. (2.13). For discrete case, the parameters a and b in Eq. (2.12) are discretized as:

$$a = a_0^m,\ b = nb_0a_0^m \tag{2.14}$$

for m, n are integer.

In Eq. (2.14), a_0 and b_0 are constants that determine the sampling intervals of a wavelet with $a_0 > 1$ and $b_0 > 0$. The discrete wavelet transform (DWT) is then defined as:

$$W_f^d(m,n) = \int_{-\infty}^{\infty} \psi_{mn}(t)f^*(t)dt \tag{2.15}$$

where

$$\psi_{mn}(t) = a_0^{\frac{-m}{2}}\psi(a_0^{-m}t - nb_0). \tag{2.16}$$

There are a number of "wavelets" for us to perform the time-frequency analysis. Among them, the dyadic wavelets are of most interest. In the studies of dyadic wavelets, a_0 and b_0 in Eq. (2.16) are chosen as $a_0 = 2$ and $b_0 = 1$, such that Eq. (2.16) is rewritten as:

$$\psi_{mn}(t) = 2^{\frac{-m}{2}}\psi(2^{-m}t - n). \tag{2.17}$$

Dyadic wavelets can be introduced using the technique of multiresolution analysis (MRA) and filter banks [50]. MRA writes a $L^2(\mathbf{R})$ function $f(t)$ as a limit of successive approximations, each of which is a smoothed version of $f(t)$, by decomposing the space $L^2(\mathbf{R})$ into a ladder of low-pass subspaces V_j that satisfy the following properties:

$$V_j \subset V_{j+1} \tag{2.18}$$

$$f(t) \in V_j \Leftrightarrow f(2t) \in V_{j+1} \tag{2.19}$$

$$\bigcap_{j \in Z} V_j = \{0\} \tag{2.20}$$

$$\bigcup_{j \in Z} V_j = L^2(\mathbf{R}) \tag{2.21}$$

$$\exists \varphi(t) \in V_0 \quad s.t. \quad \{\varphi(t-k)\}_{k \in Z} \text{ .is a Riesz basis for } V_0. \tag{2.22}$$

The function $\varphi(t)$ is usually called the scaling function. From Eq. (2.19), we have the "two-scale difference equation":

$$\varphi(t) = 2\sum_k h_k \varphi(2t-k) \tag{2.23}$$

where the sequence $h_k \in l^2(\mathbf{Z})$. The lost details from reducing resolution from V_{j+1} to V_j is defined as W_j. That is, $V_{j+1} = V_j \oplus W_j$. A function is called "wavelet" if the collection of $\{\psi(t-k)\}_{k \in Z}$ forms a Riesz basis in W_0, and, the collection of $\{\psi(2^n t-k)\}_{n,k \in Z}$ forms a Riesz basis in $L^2(\mathbf{R})$. If P_j and Q_j are defined as the projection operator onto V_j and W_j, respectively, then we have,

$$P_j f + Q_j f = P_{j+1} f \text{ and } f = \sum_j Q_j f .$$

Since W_0 complements V_0 to form V_1, the wavelet $\psi(t)$ can also be expressed in terms of $\varphi(2t)$. Thus, there exists a sequence $g_k \in l^2(\mathbf{Z})$ such that

$$\varphi(t) = 2\sum_k g_k \varphi(2t-k). \tag{2.24}$$

It can be shown from Eqs. (2.23) and (2.24) that, the DWT of a signal $f(t)$ can be done by performing appropriate low-pass and high-pass filtering using filter bank g and h, through MRA starting with the scaling function $\varphi(t)$. The actual MRA tree structured implementation using filter bank for the wavelet decomposition and reconstruction of a function with dyadic samples is shown in Fig. 2.3.

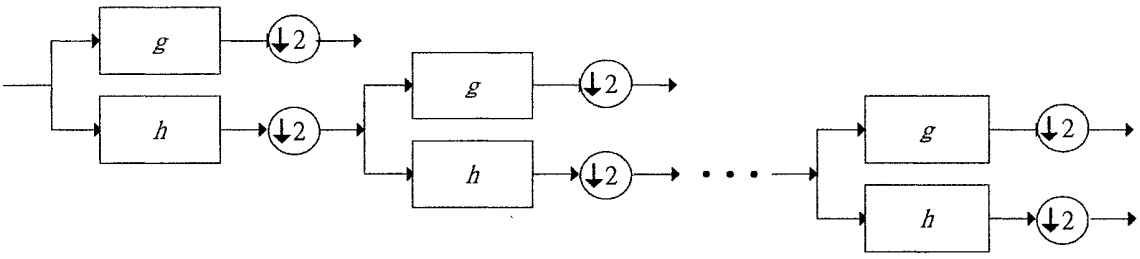


Figure 2.3: Computation of wavelet coefficients via multiresolution analysis filter bank.

In Fig. 2.3, the symbols ↓2 and ↑2 stand for subsampling and upsampling by two, respectively. Subsampling by two drops every other point in the sequence and the sequence is renumbered. Upsampling by two tries to insert a zero between every two points in a sequence and the sequence is also renumbered. Further discussion on the relationship between the continuous wavelet transform, discrete wavelet transform, multiresolution analysis as well as wavelet design can be found in [49, 50, 81].

2.3.2 Noise reduction using wavelet-based methods

The process of noise reduction using wavelet methods is often called wavelet denoising [54]. The advantage of wavelet denoising over other classical noise reduction algorithms is that discontinuity and noise-like structure can be easily characterized in the wavelet domain. Hence, some wavelet denoising techniques work by distinguishing and removing the noise components among the wavelet coefficients of the signal while retaining as much as possible the important signal features. Many approaches have been developed for wavelet denoising and particularly, denoising through wavelet-based singularity detection, thresholding and statistical image modeling are the methods that have aroused much attention.

One of the most powerful properties of the wavelet transform is its capability for detecting the local regularity of signal [51]. Singularities carry important information, such

as edges and textures, which are of particular importance in many applications. By estimating the Lipschitz exponent of a singularity, one can distinguish and remove the irregular parts, which correspond to the noise, of the signal to achieve the goal of denoising [51, 82]. In particular, Mallat proposed to estimate the Lipschitz exponent of a singularity by tracing its wavelet transform modulus maxima (WTMM) [51]. Mallat showed that the local regularities of certain types of non-isolated singularities in a signal can be characterized by their Lipschitz exponents. They also showed that the decay of the expected WTMM value of a wide sense stationary white noise across scales is inversely proportional to $1/2^j$, where j is the scale. This means, the WTMM curves of noise are expected to decay across scales at least at a rate of $1/2^j$. Sometimes, they may even not propagate to coarser scales. On the other hand, it is also known that the edges of a signal possess zero Lipschitz exponents while the regular parts of the signal possess positive Lipschitz exponents. They indicate that the values of the WTMM of them will be the same, and will not increase, when scale increases. Consequently, in the application of denoising, we can remove the noise of a signal by removing all WTMM whose amplitude decreases when scale increase or adjusting the maxima curves that possess negative Lipschitz exponent using non-decimated wavelet transform. The denoised signal is then reconstructed from the processed WTMM.

Although the WTMM based algorithms give promising performance in many aspects, the irregular sampling nature of WTMM complicates the reconstruction process. Furthermore, examples were found [83] to show that the WTMM representation cannot uniquely characterize a signal. Hence, the reconstruction of a signal from its WTMM may not be consistently stable. Consequently, many researchers suggested other methods to estimate a signal from its WTMM representation. For example, in [84] and [85], a non-expansive approach was proposed to reconstruct a signal from its WTMM representation. Besides, the discrete time wavelet extrema representation [86] was also proposed as an alternative convex representation for the WTMM to enable consistent reconstruction.

However, the proposed reconstruction methods are iterative and thus require high computational complexity. To further reduce the computational complexity, [87] suggested to compute the wavelet transform modulus sum (WTMS) over the regularly located wavelet coefficients for the selection of irregular component to avoid ambiguous operations.

In wavelet analysis, it is found that signals can be represented by a few large wavelet coefficients while white noise signals are represented by coefficients that are scatter in the wavelet domain. Therefore, the reconstruction of signal from only the large wavelet coefficients will tend to preserve most of the signal energy but little of noise energy. From the above phenomenon, Donoho [52, 53, 54] suggested another denoising technique by performing hard thresholding or shrinkage on the wavelet coefficients of a signal using orthogonal wavelet transform [49]. It is shown that due to the vanishing moment property and the compact support of the wavelet transform, most signal energy is clustered in a few wavelet coefficients. On the other hand, since noise is irregular in nature, its wavelet coefficients will be scatter on the time axis in a random manner. The thresholding/shrinkage of the wavelet coefficients with a proper threshold [88] can then significantly reduce noise and achieve near optimal performance in the minimax sense. This operation guarantees that the denoised signal is at least as smooth as the input noisy signal.

The performance of the wavelet shrinkage is expected to differ from the ideal performance [53] by at most a factor of $2\log n$, where n is the total number of samples. However, the denoised signal may contain spurious oscillations due to the translation variant property of the decimated wavelet transform. Therefore, many variants of the wavelet thresholding/shrinkage techniques were developed. For example, in [89], a cycle spinning method that averages the denoised results from all the possible shifts of the noisy signal was proposed to reduce the effect of spurious oscillations. In [90], undecimated wavelet basis for denoising was also proposed for the same purpose. Furthermore, another class of wavelet basis, which possesses better shift invariant property, was introduced in

[91] to improve the performance of wavelet shrinkage technique for denoising. Nevertheless, all of these approaches do not guarantee that edges can be preserved in denoising.

Another fast expanding class of wavelet-based methods for denoising adopts the techniques of statistical image modeling [92, 93] in the wavelet domain. Actually, there are often significant correlations between the wavelet coefficients in the transforms of real images. For example, it is observed that large coefficient values cascade along the branches of the wavelet tree. Statistical image modeling techniques, such as Hidden Markov Tree (HMT) [92], are then proposed to capture the key features of the joint statistics of the wavelet coefficients and allow effective denoising than other methods which assume wavelet coefficients to be independent [93]. This is achieved by means of hidden state variables that describe the likely characteristics, such as the amplitudes, of the wavelet coefficients. A Markov model is used to describe the relative probability of transitions between different states along a branch of the wavelet tree moving from coarse to fine scales. Nevertheless, the proposed techniques in this field require high computation complexity due to their iterative estimation processes.

2.4 Error Resilient Image and Video Communications

When transmitting compressed image and video bitstreams through a communication channel, channel errors, caused by different loss mechanisms, can corrupt the data and severely degrade the reconstructed picture quality at the receiver end. The effects of errors are very often magnified by the VLC coding techniques adopted in most current compression schemes. For video communications, the effects of errors may further propagate, due to the use of predictive coding in video compression, in temporal direction.

Therefore, one of the crucial issues for robust image/video communications is error resilience. In this section, we overview the fundamental techniques related to the error resilient image and video coding. Firstly, current transform coding schemes for image/video compression are briefly discussed. Secondly, type of channel errors and their effects to compressed image/video bitstream are presented. Subsequently, methods to detect error in a corrupted bitstream are studied. Finally, some error resilient coding techniques, without the use of feedback channel, are elaborated.

2.4.1 Transform coding for image/video compression

Transform coding has grown to be the dominant approach for nowadays image and video compression. Fig. 2.4 shows a typical transform coding system for most image/video source encoders. In Fig. 2.4, the input picture is either a still image or a motion-compensated video frame. As illustrated in Fig. 2.4, the input pictures are first transformed to either the DCT or wavelet domains followed by a quantization process. Afterward, the quantized coefficients are converted into symbols according to some data representation formats. Finally, the output symbols are then encoded using entropy coding techniques.

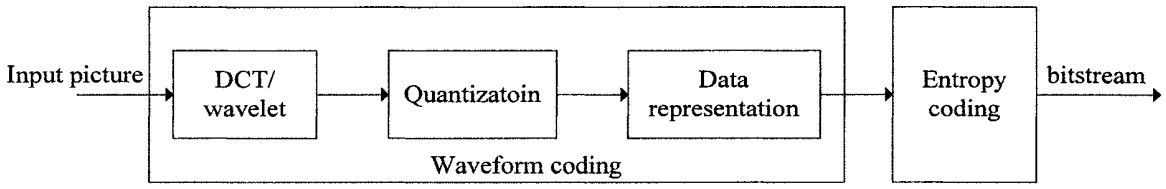


Figure 2.4: Typical transform coding system in image/video source encoder.

As shown in Fig. 2.4, the combination of transformation, quantization, and the conversion to specific data representation is known as the waveform coding process [94]. This waveform coding process tries to remove the statistical and psychovisual redundancies

presented in the input picture while the entropy coding aims at removing coding redundancies. There are a number of transform coding schemes proposed in the literature. Based on the difference in the adaptation of transformation tools, previously proposed transform coding schemes can be classified into two categories, namely: DCT-based coding, and wavelet-based coding. In these coding schemes, compressions are achieved by reducing the correlation among the input images/videos. The basic operations for these schemes are described in the following sections.

A. DCT-based coding

DCT-based coding scheme tries to divide the input picture source into a number of 8×8 blocks. Each block is transformed into the DCT domain followed by a perceptual quantization process. For most images, adjacent pixels are highly correlated. DCT lays the foundation for achieving data compression by exploring the correlation among pixels and concentrating most of the signal energy in the lower spatial frequencies of a block. Usually, all quantized DCT coefficients, in each block, are zig-zag scanned and rearranged into a 1-D array. This ordering helps to facilitate entropy encoding by placing low-frequency non-zero coefficients before high-frequency coefficients. Entropy coding, such as Huffman coding, is finally applied to the coefficients in order to achieve additional compression by exploring the statistical characteristics of the coefficients. One example of DCT-based image coding schemes is the well-known JPEG [7] still image compression standard.

DCT-based scheme is also applied to video coding. It is well-known that high temporal correlation exists between neighboring frames in a video sequence. For this reason, the motion compensated differential pulse code modulation (MC-DPCM) [95] is often applied and generates coding coefficients with small magnitude that is suitable to the later part of the coding system. Such coding coefficients are then encoded using the DCT-

based coding scheme to achieve further improvement in coding efficiency. H.263, MPEG-1, MPEG-2, and MPEG-4 are the well-known examples of video codec that adopted the DCT coding scheme.

B. Wavelet-based coding

All wavelet-based image coding schemes convert input picture from the spatial domain into the wavelet domain using DWT with multiple level dyadic decomposition. In most cases, a uniform quantization is applied but the data representation method adopted for different approach is quite different. For example, in EZW and MPEG-4 still texture image coding [13], all quantized wavelet coefficients are converted into several zero-tree symbols by exploring the inter-scale correlation among coefficients in order to predict the insignificant wavelet coefficients. However, in JPEG-2000, the embedded block coding with optimized truncation (EBCOT) algorithm [96] is used instead. Rather than generating a single scalable bitstream to represent the entire image, EBCOT partitions each subband into relatively small blocks of samples and generates a separate highly scalable bitstream to represent each so-called code-block. The algorithm exhibits state-of-the-art compression performance while producing a bitstream with an unprecedented feature set, including the resolution and SNR scalability together with a random access property. Entropy coding for most wavelet-based image coding schemes is achieved by means of arithmetic coding [97].

Although each subband is divided into code blocks, their coding is not performed independently but follows a RD mechanism. That is why the wavelet coding schemes can avoid blocking artifacts even at high compression rate. Wavelet-based coding is more robust under transmission and decoding errors, and also facilitates progressive transmission of images.

2.4.2 Types of error and their effect to compressed bitstream

As mentioned in Section 1.2, channel errors can be roughly classified into two main categories: random bit errors and erasure errors. Random bit errors are usually caused by the imperfection of physical channels and appear in the form of bit inversion, insertion, and deletion. They are usually quantified using a parameter called bit error rate (BER) which denotes the average probability that a bit is in error. Erasure errors are caused by packet loss in packet network, burst errors in the storage media due to physical defects, or system failures for a short period of time. They occur in the form of a loss or damage to contiguous segments of bits and can be quantified using parameters like the number of bursts, the length of a burst, and the BER within a burst. A single random bit error can also cause serious erasure errors due to the loss of synchronization in the decoder.

The effects of random bit errors and erasure errors to the image/video bitstream highly depend on the coding method and the content of the source. Both types of error can cause isolated effects or spatial error propagation within an image or video frame, and severe temporal error propagation to video sequence as discussed in the following paragraphs.

A. Isolated effects

The effect of errors may sometimes convert a VLC codeword to another valid VLC of the same length. If this error does not affect the decoding synchronization of the following bits and the damaged codeword is not a prediction for another codeword, its effect will be limited to within the corrupted codeword. Depending on the type of information carried by the corrupted codeword, impact of isolated errors can range from negligible to objectionable in an image or video bitstream. For example, if the corrupted codeword carries merely information about DCT coefficients, insignificant effect will be resulted. However, isolated errors in vital header information can cause severe distortion to the reconstruction.

B. Spatial error propagation

Spatial error propagation is usually due to either the corruption of interdependently coded data or the loss of synchronization in the decoder.

Upon the loss or corruption of interdependently coded data, current decodable codewords often provide incorrect decoded information for most predictive coding schemes. Errors will propagation from the original corrupted point to the current decoded point and then to the next predicted points following the strategy of the prediction algorithm. Generic example can be found when predictive coding of DC coefficients for a set of transformed blocks is applied. In this case, if the codeword carrying the information of DC coefficient of the first block is corrupted, then all the upcoming blocks will not be decoded correctly.

Loss of synchronization in decoder can also cause spatial error propagation. A single error can cause the decoder to falsely decode a codeword and lead to the loss of synchronization between the encoder and the decoder. This loss of synchronization expands the erroneous area spatially until a valid resynchronization marker (RM) appears. In the worst cases, even when synchronization is regained, the decoder may decode coefficients that have no meaning without the corrupted coefficients. For example, in run-length encoding, if an incorrect run-length has been decoded, all of its following data will be misplaced and error will propagate to the end of the block or even to other blocks.

C. Temporal error propagation

Propagation of errors across video frame is known as temporal error propagation. It is caused by the uses of motion estimation and compensation between adjacent frames in most video transform coding schemes. Due to the motion estimation and compensation processes, current frame is decoded in addition to the information of its reference frame(s). Therefore, errors can easily propagate from the reference frame to the current frame.

2.4.3 Methods to detect error in a corrupted bitstream

Precise error detection is a vanguard of many error resilient coding schemes. For most of the error resilient coding schemes, encoders try to strengthen the bitstream against errors on the one hand while decoders perform error concealment, upon the detection of errors, on the other hand. In general, error detection can be achieved by exploring the characteristics of the received bitstream or the reconstructed picture in decoders.

A widely used error detection method is the detection of syntax and semantic violation in the received bitstream, for example, invalid VLC codewords or header information. Another advanced method tries to detect violations according to the general characteristics of natural video signals, for example, the detection of strong discontinuities at the borders of blocks, or blocks with highly saturated colors [94]. However, this method may sometimes classify an error-free block as an erroneous one.

The performance of error concealment is heavily depended on the accuracy of the error detection. Nonetheless, due to feature of VLCs, the detected error position is usually not the actual error position in a corrupted bitstream. In this case, error concealment (or other appropriate techniques) cannot be taken immediately and hence errors are allowed to propagate spatially or temporally, as discussed in the previous section. There is no guarantee for any error detection strategy to detect all errors in a corrupted bitstream. In particular, it is usually more difficult to detect isolated errors in a corrupted bitstream. Furthermore, error detection can also be done by applying sorts of channel coding techniques, such as the FEC. However, details of channel coding will not be discussed in this dissertation.

2.4.4 Existing error resilient tools

Many resilience mechanisms have been proposed to resist the effects of errors to the image and video bitstreams and they can be categorized into four main classes, as already discussed in Section 1.3.3. In fact, channel coding also plays a very important role for error resilience in most cases. For example, FEC and interleave coding are two classical methods in this class for error resilience proposes. However, techniques of channel coding will not be discussed here. Details of these methods can be found in [17], [18] and [19]. In this section, we give a brief overview of some well-known non-interactive error resilient coding techniques.

A. Methods carried out by mainly the source encoder

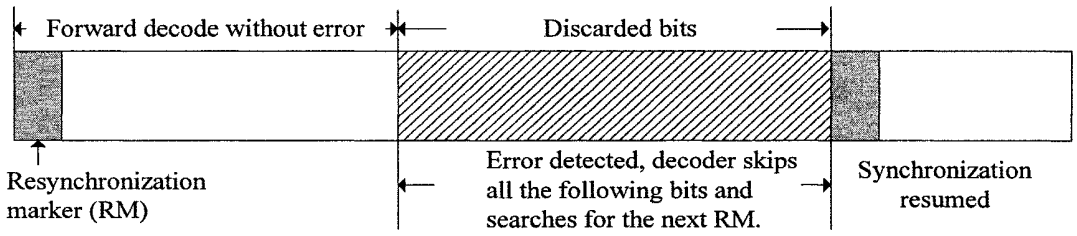
In this class of methods, the source encoder plays the primary role to add a controlled amount of redundancies to the bitstream with the drawback of reducing coding efficiency. Some techniques in this class aim at minimizing the effects of errors. Some of them are designed to facilitate error detection and concealment at the decoder. The others try to guarantee a basic level of quality while providing graceful degradation in the presence of transmission errors.

i. Use of resynchronization marker (RM)

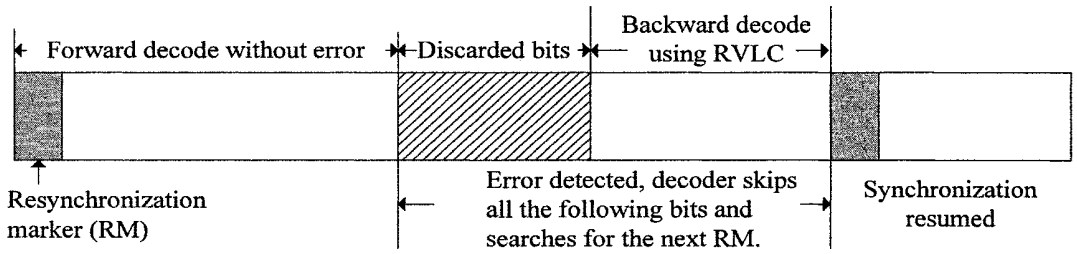
Insertion of unique resynchronization marker (RM) in a bitstream can help decoders to regain synchronization in case of error and to reduce spatial error propagation. The operation of RM is straightforward. As illustrated in Fig. 2.5(a), in case of any error found, the decoder skips all the remaining bits starting from the detected error position until a valid RM appears. The receipt of a valid RM indicates a correct decoding point for decoders to decode the rest of the following codewords and hence regain the synchronization between the encoder and the decoder.

Classical approaches often insert RMs in the spatial domain regularly. For example, in version 1 of H.263, fixed RMs (FRM) called PSC and GBSC are inserted at the starting positions of each picture and GOB layer [11]. However, this FRM approach covers a fixed number of macroblock (MB) but allows the variation of number of bits between two valid RMs. This results in an uneven distribution of RMs (in the sense of bit count) among a bitstream and makes those MBs with high activities more susceptible to errors. To deal with this problem, another packet-based approach [13], adopted in the MPEG-4 standard, tries to insert RMs in an even way throughout a bitstream such that the number of bits involved between two RMs are approximately equal.

However, in some cases, errors may corrupt or emulate the occurrence of RM. To this problem, MPEG-4 suggests the fixed-interval synchronization approach [13] to place RMs merely at legal fixed interval locations in a bitstream. In this approach, only RMs which are obtained at those legal locations will be adopted for resynchronization.



(a)



(b)

Figure 2.5: Resynchronization using (a) resynchronization marker (RM) and (b) reversible variable-length code together with RM.

ii. Use of reversible variable-length code (RVLC)

Reversible variable-length codes (RVC) are proposed mainly for accurate error detection. RVLC can be decoded in both the forward and backward directions. Its operation is similar to RMs but allows backward error tracking to localize errors in a corrupted segment. As shown in Fig. 2.5(b), when any error is detected in the bitstream, the decoder searches for the next resynchronization point. Instead of discarding all the bits between the detected error position and the next resynchronization point, the decoder attempts to decode the corrupted segment in a reverse manner in order to minimize the possible corrupted segment. This not only enhances the accuracy of detected error position but also reduces the number of uncorrupted bits to be discarded.

iii. Adding redundant information

Apart from using RMs, adding supplementary information or repeating some important information in a bitstream can also assist error handling at decoders. For example, In MPEG-2 standard [98], the encoder can optionally send auxiliary motion vectors, with respect to a reference frame, for intra-coded MBs. These motion vectors can be used for effective error concealment of the corresponding MBs in case of error. Similar technique, which is known as the header extension code (HEC) [13], suggested by the MPEG-4 standard, tries to repeat important header information among the bitstream in order to support possible error concealment.

iv. Restricted prediction

The main cause of spatial and temporal error propagation is the use of predictive coding in most source encoders. Therefore, one of the most effective ways to reduce spatial or temporal error propagation is to limit or carefully select the range for prediction. It can be done by dividing each video frame into several segments (as recommended in the Annex R) [11] of H.263, and encoding each segment without referring to other segments in the same frame and the reference frame(s). By doing so, errors occur in a particular segment will only affect the data within that segment but not propagate to other segments.

v. Layered coding (LC) with prioritization

Layered coding (LC), which is sometimes called scalable coding, tries to guarantee a basic level of quality while providing graceful degradation in case of error. Similar to the segment-based coding as mentioned above, in layered coding, video is divided into a base layer and one or more enhancement layers. The base layer is separately encoded and decodable. Base layer provides a basic level of perceived quality while the enhancement layers can be decoded to improve the reconstruction quality at the decoders. In the

application of error resilience, prioritized transmission, which is sometimes called unequal error protection (UEP), is usually combined with layered coding.

In most cases, the base layer is transmitted with higher priority or a higher degree of error protection to guarantee certain level of reconstructed picture quality. For example, in the data partitioning approach in the MPEG-4 standard, coded data in each MB are divided into two portions with different priorities. Those important information, such as motion vectors and low frequency data, are placed in the high priority layer while less important data, such as the high frequency coefficients, are placed in the low priority layer. The problem of LC is that a loss in the base or high priority layer can lead to a disastrous effect in the decoded visual quality.

There are a number of approaches to divide video into more than one layers by performing spatial, temporal, and SNR scalability codings [94]. This class of techniques achieves the goal of error resilience by considering both the source and channel coding. Therefore, it can be classified as belonging to the joint source and channel coding, as discussed in Section 1.3.3.

vi. Multiple description coding (MDC)

An alternative approach to combat transmission errors is by using multiple description coding (MDC). With this coding scheme, several bitstreams, which are referred as descriptions, of the same source signal are generated and transmitted over separate channels. The MDC encoders and decoders are designed such that the quality of reconstructed signal is acceptable with any one description, and that incremental improvement is achievable with additional descriptions. This technique also assumes that there are multiple channels between the encoder and the decoder. These multiple channels can be physically distinct paths or they can be a single path divided into multiple virtual channels using, for example, time or frequency division. The technique further assumes that

the error events of these multiple channels are independent. This means that the probability that all channels simultaneously experience errors is very small.

By assuming the independence of error in the channels, it is cleared that at least one description will be received at the decoder without errors. This description carries some fundamental information about the transmitted video and can, therefore, be used to provide a basic level of quality. As all descriptions share some fundamental information about the source, so they are correlated. Hence, missing descriptions can be estimated from correctly received descriptions such that reconstruction quality can be improved. An advantage of MDC over LC is that it does not require special provisions in the network to provide a reliable subchannel. This MDC techniques can be referred as one of the joint source and channel coding techniques also.

vii. Intra-refreshing

Another effective way to stop error propagation is to intra-code some of the video frames or MBs according to certain schedules. This technique is often referred as intra-refreshing. Intra-refreshing stops both the spatial and the temporal error propagation due to prediction and motion estimation/compensation. However, intra-refreshing a video frame significantly decreases the coding efficiency. Performing intra-refresh at the MB level would be a more appropriate choice than at the frame level. In general, the required number of MBs to be refreshed highly depends on the channel quality and capacity as well as the video content.

There are many schemes [12, 58, 99, 100, 101] proposed to determine the spatial location and the number of intra-coded MBs. In particular, the random intra-updating (RIU) [12] technique, adopted by the H.263+, tries to divide MBs into individual groups randomly and perform intra-refreshing among the groups periodically. This RIU technique is simple and easy to implement. However, it neglects the content of the video such that extra bit

resource is used to intra-refresh low activity regions which can be recovered effectively by performing appropriate error concealment.

Another class of approaches is to perform intra-refreshing based on RD optimization [99, 100, 101] by considering the effects of error propagation and concealment. Obviously, if there is a feedback channel between the encoder and the decoder, information regarding the number and the locations of damaged MBs can help the encoder to better decide the number and the locations of intra-coded MBs.

viii. Error resilience entropy code (EREC)

An interesting alternative to inserting redundancy is the error resilience entropy coding (EREC). The basic operation of the EREC is to rearrange variable-length blocks of data into fixed-length slots in such a way that the decoder can independently find the start of each block and start decoding it. For example, if there are N variable-length blocks with lengths b_i, for i = 1, 2, ..., N, the encoder first chooses a total data size $T = \sum b_i$, which is sufficient to encode all the data. This total data size is split into N slots of fixed lengths, s_i, for i = 1, 2, ..., N. An N-stage algorithm is then used to place the data from the variable-length blocks into the fixed-length slots. At each stage, a block i with data left unplaced searches slot $j = i + \phi_n \pmod N$ for space to place some or all of the remaining data as described in [60, 61] until all the bits are packed into one of the slots. ϕ_n is an offset sequence that is usually pseudo-random.

For a fixed size of slot, the decoder can regain synchronization at the start of slot without the need of RM in the case of losing synchronization. Besides, concept of LC can also be applied by placing less important data in later stages of the algorithm.

B. Post-processing methods executed at the decoder to conceal errors

This class of methods are often called as error concealment in which decoders play the primary role to hide the effects of errors by providing a subjectively acceptable approximation to the original data without any interaction with the encoder. This is achieved by exploiting the limitation of the human visual system and the high temporal and/or spatial correlation of most images and videos. Depending on the information used for hiding the errors, error concealment techniques can be divided into three main categories: spatial techniques, temporal techniques, and hybrid techniques.

i. Spatial error concealment

Spatial techniques exploit the high spatial correlation of image and video signals and conceal damaged content in a frame using information from correctly received and/or previously concealed neighboring pixels within the same frame through the means of interpolation.

Recovering of loss image data can be achieved by performing interpolations in either the spatial or the transformed domains. For example, in [69], a spatial interpolation algorithm based on the projections onto convex sets (POCS) theory was proposed. This method makes use of spatial information only in the reconstruction process such that it is applicable to both the intra-coded blocks and still images. For inter-coded blocks in block-based video coding, the motion information has to be recovered first before the POCS method can be applied.

On the other hand, a frequency domain interpolation algorithm was proposed in [102] to predict lost coefficient in a damaged block from the corresponding coefficients in its four neighbor blocks. A more precise approach proposed in [103] suggested to interpolate pixel values within a damaged MB from its four 1-pixel-wide boundaries to improve the estimation accuracy.

ii. Temporal error concealment

Temporal techniques exploit the high temporal correlation of video signals and conceal damaged content in a frame using information from correctly received and/or previously concealed pixels within a reference frame. However, these techniques will completely fail in cases like scene changes and uncovered background. As in motion-compensated prediction, the process of temporal concealment involves two stages [94]: concealment displacement estimation and displacement compensation. The concealment displacement estimation stage aims at recovering motion information while the displacement compensation tries to approximate the original data by motion compensation using the corresponding motion information. In fact, the only difference between the various conventional temporal techniques reported in the literature is in their displacement estimation algorithms.

In displacement estimation, the simplest technique is to set the damaged motion vector to zero based on the center-biased property of video block-motion fields and directly replace the damaged block by copying the corresponding block in the reference frame. This method is generally known as temporal replacement [104] and it works well for stationary and quasi-stationary areas but will fail for fast-moving areas. Another approach tries to replace the damaged motion vector with average or median [104] of neighboring vectors. This technique works well for area with smooth motion. A boundary matching (BM) [105] and Bayesian statistical [106] approaches have also been investigated for motion vector recovery.

iii. Hybrid Error Concealment

Hybrid techniques exploit both spatial and temporal correlations of video signals. A straightforward technique is to use spatial concealment for intra-coded blocks and temporal

concealment for inter-coded blocks. However, the coding mode information of a damaged block is very often corrupted. Thus, coding mode information needs to be recovered first before appropriate displacement compensation can be applied.

Usually, there is a high correlation between the coding modes of adjacent blocks. Thus, one way to estimate the coding mode of a damaged MB is by collecting the statistics of the coding mode pattern of its adjacent MBs, and find the most likely mode given the modes of surrounding MBs [107].

2.5 Summary

In this chapter, we have briefly reviewed the existing approaches which are relevant to the present work. In particular, the background of image restoration and a summary of blind restoration of blurred image using the techniques of ARMA modeling and GCV parameter estimation are presented. We also discuss the wavelet theory and its applications to image noise reduction. Moreover, a brief overview of the aspects related to error resilient image/video communications is provided. Techniques for current image/video transform coding, types of channel errors and their effects to compressed image/video bitstream, methods to detect error in a corrupted bitsream, and a review of the current error resilient tools for source coding are discussed.

In the following chapters, novel techniques are introduced to the problems of blurred image restoration, wavelet based noisy image compression, and efficient error resilient source coding, accordingly.

Chapter 3 Blind Restoration of Blurred Image using Discrete Periodic Radon Transform (DPRT)

3.1 Introduction

As mentioned in the previous chapters, the imperfection of the image formation process often blurs and degrades the acquired image or video frame. By assuming that the blurring process is a convolution process between the blur function and the real source, the image formation process can often be formulated by a linear model as Eq. (2.1) [28]:

$$g(x,y) = h(x,y) ** f(x,y) + n(x,y) \tag{3.1}$$

where ** represents the 2-D linear convolution operator, $f(x,y)$ is the original image, $g(x,y)$ is the observed image, $h(x,y)$ is the point spread function (PSF) and $n(x,y)$ is the additive noise due to the imaging system, respectively. The problem of image restoration is to recover an unknown original image $f(x,y)$ from a given blurred image $g(x,y)$ and the PSF $h(x,y)$. Different approaches have been proposed [108, 109] to solve this problem and they have shown to have good performance when priori information of the original image is acquired. Unfortunately, these approaches rely on the assumption that full information of the PSF is available can hardly be fulfilled in many practical situations. Without the prior information of the PSF, the restoration problem becomes the so-called blind image restoration, which can be tackled as a single problem or divided into two steps, that is, "blindly" identifying the PSF and then performing the image restoration [31, 32].

One of the important approaches for solving the blind restoration problem is based on the ARMA parameter estimation method as described in Section 2.2.1. By identifying the ARMA parameters, we can identify the true image and PSF. The maximum-likelihood (ML) estimation [34, 35, 36] and generalized cross-validation (GCV) approaches [37, 38] are the two most popular approaches to estimate the ARMA parameters. The ML approach has a long history for ARMA parameter estimation. It accomplishes the estimation by seeking for the parameters that produce the blurred image on hand. The GCV approach on the other hand achieves the estimation by looking for the parameters that minimize a weighted sum of prediction errors. This criterion has been shown to possess certain properties that are superior to those of ML in the regularization parameter estimation [42] and perform much better for real images [38, 110]. However, the GCV approach is often criticized due to extensive numerical searches for minimizing the GCV score. In [111], efficient approach was proposed to first estimate the unknown PSF parameter set and regularization parameter from raw data, and then make use of a computationally inexpensive algorithm to solve the non-blind problem. However, this approach requires multiple blurred versions (or low-resolution) of the original image during restoration (or super-resolution) that imposes limitation to general applications.

In this chapter, we consider adopting the discrete periodic Radon transform (DPRT) [74, 75] to estimate the PSF. The recently proposed DPRT has many properties similar to the conventional continuous Radon transform [112]. They include the discrete Fourier slice theorem and the circular convolution property [75]. Using these properties, a 2-D signal can be processed by some 1-D approaches to reduce the computational complexity. DPRT is different from other discrete Radon transform [113] in that it has an exact inverse transform. In the following sections, we first review the concept of DPRT, its properties and inversion formula as well as the technique to use GCV for image restoration. Then, we elaborate how to convert the 2-D GCV to the 1-D ones using DPRT for blind image restoration. Finally,

systematic study is carried out to show that the new approach can be an order of magnitude faster than the traditional approach while keeping similar quality restored images. Besides, the memory requirement is also dramatically reduced.

Since the proposed approach is just a fast algorithm to the original 2-D GCV algorithm, it shares the same advantages and limitations in applications. For instance, the GCV methods are usually less sensitive to additive noise than methods of other classes of blind deconvolution techniques because they take into account the noise in the system [31]. However, since the GCV methods deal only with the second-order signal statistics, phase cannot be identified uniquely in the restoration procedure. Hence usually the GCV approaches are used for restoring images blurred by linear motion or the out-of-focus PSFs, which are symmetric and zero phase in nature.

3.2 Review on Discrete Periodic Radon Transform (DPRT)

The discrete Radon transform (DRT), which is considered to be the discretization of the continuous Radon transform [112], has been generalized [113] and adopted in widespread applications, such as computer tomography, magnetic resonance imaging, remote sensing, and geophysics, etc [112]. However, the inversion of DRT, which is necessary in these applications to retrieve the original image, often poses problems in realization. For example, it may require excessive arithmetic operations to interpolate an image from a polar to a rectangular grid. Hence, a number of alternative approaches were proposed. In particular, Beylkin [113] proposed an exact inversion algorithm for DRT. He showed that if the discrete version is based upon Radon's original formula, reconstruction can only be an approximation. Based on the approaches given by Beylkin, Kelly and Madisetti [114] further proposed the fast Radon transform (FRT) and the inverse fast Radon transform

(IFRT) which eliminate the interpolation calculation and convert directly between a raster scan grid and rectangular/polar grid in one step. Nevertheless, these approaches require the implementation of a number of forward and inverse FFTs and the problem of excessive arithmetic operations still exists. Another variant of DRT has also been proposed in [115]. However, the DRT proposed in [115] can only serve as an alternate means for the computation of 2-D discrete Fourier transforms. No investigation on its inversion or properties was performed. Hence, it cannot be considered as a transform in the strict sense.

On the other hand, the finite Radon transform has been used in combinatorial [116] and image representations [117]. This transform of a real function f, defined on a finite set S, is obtained by summation over a certain set of blocks. A scheme of finite Radon transform on groups and a thorough study of the finite Radon transform on image with size $P\times P$, where P is prime, were introduced in [117]. Although that approach has many advantages as compared to the earlier versions of DRT, the strict requirement on the image size ($P\times P$) has imposed much difficulty in actual implementation. To this problem, a new variant of the DRT, called the discrete periodic Radon transform (DPRT), was proposed.

DPRT was first proposed by Lun *et al* in 1995 and it preserves most of the useful properties of the continuous Radon transform. For example, the Fourier slice theorem and the convolution property. As compared to the previous DRT, the DPRT has several advantages. Firstly, it is an exact transform; secondly, it can work on images with size $P\times P$ or $2^n\times 2^n$, where n is any positive integer; thirdly, the inverse DPRT is multiplication free such that the finite word-length problem in actual implementation can be eliminated.

Denote the set of real number as $\mathbf{R}$, $\mathbf{R}\times\mathbf{R}$ as $\mathbf{R}^2$, the set of integers as $\mathbf{Z}$, $\mathbf{Z}\times\mathbf{Z}$ as $\mathbf{Z}^2$, a subset of integers $\{0,1,2,...,n-1\}$ as $\mathbf{Z}_n$, {0,2,4,...,n-2} as $\dot{\mathbf{Z}}_n$ and $\mathbf{Z}_n\times\mathbf{Z}_n$ as $\mathbf{Z}_n^2$. Denote $l^2(\mathbf{Z}_n^2)$ to be a set of doubly sumable 2-D complex functions over $\mathbf{Z}_n^2$ and $L^2(\mathbf{Z}_n^2)$ to be a set of doubly integrable 2-D complex function over $\mathbf{R}^2$. Let $f(x,y)$ be a 2-D function with $x,y\in\mathbf{Z}$.

The discrete periodic Radon transforms (DPRT) of $f(x,y)$ on $\mathbf{Z}_P^2$ and $\mathbf{Z}_{2^n}^2$ were proposed in [75] as:

Case 1: The DPRT of $f(x,y) \in l^2(\mathbf{Z}_P^2)$: $l^2(\mathbf{Z}_P^2) \rightarrow l^2(\mathbf{Z}_P^2 \cup \mathbf{Z}_P)$ is defined as:

$$f_m^c(d) = \sum_{x=0}^{P-1} f(x, < d + mx >_P); \tag{3.2a}$$

$$f_0^b(d) = \sum_{y=0}^{P-1} f(d, y) \tag{3.2b}$$

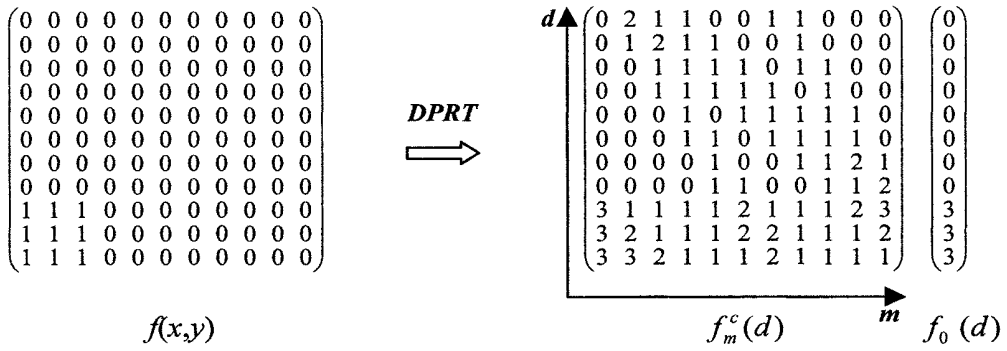
where $<A>_D$ refers to the residue of A modulo D; $d, m \in \mathbf{Z}_P$.

Case 2: The DPRT of $f(x,y) \in l^2(\mathbf{Z}_{2^n}^2)$: $l^2(\mathbf{Z}_{2^n}^2) \rightarrow l^2(\mathbf{Z}_{2^n}^2 \times (\mathbf{Z}_{2^n}^2 \cup \dot{\mathbf{Z}}_n))$ is defined as:

$$f_m^c(d) = \sum_{x=0}^{2^n-1} f(x, < d + mx >_{2^n}); \tag{3.3a}$$

$$f_s^b(d) = \sum_{y=0}^{2^{n-1}-1} f(< d + 2sy >_{2^n}, y) \tag{3.3b}$$

where $d,m \in \mathbf{Z}_n$; $s \in \dot{\mathbf{Z}}_n$. Both $f_s^b(d)$ and $f_m^c(d)$ represent the summations of the discretized contents in $f(x,y)$ with parameters $\{m,s\}$ and d. To draw an analogy with the conventional Radon transform, we may consider $f_m^c(d)$ and $f_s^b(d)$ as the projections of $f(x,y)$ in different angles "m" and "s" and different distances "d" from the origin. We shall use the term "projection" in the later part of this dissertation to indicate the DPRT of a function with a particular m or s. Fig. 3.1 shows the DPRT of a function f with support size 3×3:



(a) Original 2-D function (b) The DPRT of $f(x,y)$

Figure 3.1: An example of the prime-length DPRT.

DPRT has a few important properties, such as the discrete Fourier slice theorem and the convolution property [75], which are particularly useful in image processing. With the convolution property, a 2-D circular convolution can be decomposed into some 1-D circular convolutions. More specifically, if $g(x,y)$, $f(x,y)$, and $h(x,y)$ are all 2-D functions with supports $P \times P$ such that:

$$g(x,y) = f(x,y) \otimes_2 h(x,y)$$

where $\otimes_2$ stands for 2-D cyclic convolution. It can be shown that, after DPRT,

$$g_m^c(d) = h_m^c(d) \otimes f_m^c(d) \tag{3.4}$$

$$g_0^b(d) = h_0^b(d) \otimes f_0^b(d) \tag{3.5}$$

where $m \in \mathbf{Z}_P$. The symbol $\otimes$ stands for 1-D circular convolution. The functions $\{g_m^c(d), g_0^b(d)\}$, $(h_m^c(d), h_0^b(d))$, and $\{f_m^c(d), f_0^b(d)\}$ are the DPRTs of $g(x,y)$, $h(x,y)$, and $f(x,y)$, respectively. By performing the inverse DPRT on $g_m^c(d)$ and $g_0^b(d)$, the convolution result $g(x,y)$ can be obtained. The inversion of DPRT on $\mathbf{Z}_p^2$ is given by:

$$f(x,y)=\frac{1}{P}\left[\sum_{m=0}^{P-1} f_m^c\left(< y-mx >_P\right)-\sum_{d=0}^{P-1} f_0^b(d)+f_0^b(x)\right]. \tag{3.6}$$

And the inversion of DPRT on $\mathbf{Z}_{2^n}^2$ can be formulated as:

$$f(x,y)=S_0(x,y)-\sum_{i=1}^{\log_2 N-1} S_i(x,y)-\frac{2}{N^2}\sum_{d=0}^{N-1} f_0^c(d) \tag{3.7a}$$

where

$$S_i(x,y)=\frac{1}{2^i N}\left[\sum_{m=0}^{N/2^i-1}\sum_{r=0}^{2^i-1} f_m^c(< y-mx >_{N/2^i}+\frac{rN}{2^i}) + \sum_{s=0}^{N/2^{i+1}-1}\sum_{r=0}^{2^i-1} f_s^b(< x-2sy >_{N/2^i}+\frac{rN}{2^i})\right]. \tag{3.7b}$$

It is seen in Eqs.(3.6) and (3.7) that the original image $f(x,y)$ is reconstructed from its DPRT using mainly additions (except the scaling factors). Indeed, both the forward and the inverse DPRT require only additions for their implementations. This feature not only reduces the computational complexity but also eliminates the finite word-length error that may be introduced if multiplication is required. It is particularly useful to some computing environments where the complexity of implementing multiplications is much higher than additions. Detailed description of the properties and the inversions of DPRT can be found in [75]. Nevertheless, DPRT also has some undesirable features that make it difficult to apply in some image processing applications. As shown in Fig. 3.1, the basic version of DPRT is not orthogonal (although an orthogonal version is deemed to be available [118]). Redundancy is introduced after the transformation hence DPRT is basically not suitable to image compression. Secondly, DPRT of a compact-supported function is not necessarily compact support, as shown in Fig. 3.1. We shall show that this undesirable feature introduces much difficulty in blind image restoration problems.

3.3 Generalized Cross-validation (GCV) for Image Restoration

The GCV image restoration algorithm assumes that the image formation and the blurring processes follow the ARMA model as discussed in Section 2.2. It has been shown from Eq. (3.1) that the model of degraded images can be viewed as a 2-D MA process in the presence of noise. By recalling Eq. (2.8), the image formation process can also be modeled as a 2-D AR process:

$$f(p,q) = \sum_{(k,l)\in R_a} a_{kl} f(p-k,q-l) + u(p,q) \tag{3.8}$$

where $f(p,q)$ is the pixel value of the original image at (p,q); $u(p,q)$ is an independent zero-mean white noise with variance σ_u^2; R_a is the non-symmetric half-plane (NSHP) support [119] of the AR process; and a_{kl} is the AR formation model parameter, respectively. Eqs.(3.1) and (3.8) can be further expressed in compact form as:

$$f = \bar{a}f + u \tag{3.9a}$$

$$g = \bar{h}f + n \tag{3.9b}$$

where the 2-D signals have been lexicographically ordered. h and a are expressed in matrix-vector notations as $\bar{h}$ and $\bar{a}$. Eqs.(3.9a) and (3.9b) can be combined to form a single equation:

$$g = \bar{h}(I-\bar{a})^{-1}u + n. \tag{3.10}$$

The blur identification becomes a matter of determining the parameters of an ARMA model.

Cross-validation is well recognized in data analysis. It is also known as "leave-one-out" [120] or predictive sample reuse [121]. As discussed in Section 2.2.2, the theory of cross-validation in image restoration is simple. For a fixed model parameter set, a restored image is determined using all but one of the values from the observed image. The restored

image is then reblurred to predict the observation that was left out of the restoration. The process is iterated, using different sets of parameters each time, until all values from the observed image have been exhausted. The parameter set that minimizes the mean-square prediction error over all the observations is chosen as the optimal estimate. Due to the difficulty in implementation, the cross-validation approach is modified to become the generalized cross-validation, which is a rotation invariant form of the ordinary cross-validation. The parameter set minimizing the following GCV score is taken to be the final solution:

$$V(\theta) = N^2 \frac{\sum_{i=0}^{N^2-1} \left[\frac{\alpha}{|\gamma_i(\theta)|^2 + \alpha} \right]^2 |\hat{g}_i|^2}{\left[\sum_{i=0}^{N^2-1} \frac{\alpha}{|\gamma_i(\theta)|^2 + \alpha} \right]^2} \tag{3.11}$$

where $\theta = \{\alpha, \bar{h}, \bar{a}\}$ is the parameter set to be estimated. $\bar{h}$ and $\bar{a}$ are the PSF and image model parameters; α is the regularization parameter, respectively. The functions $\gamma_i(\theta)$ and $\tilde{g}_i$ in Eq. (3.11) relates to $\bar{h}$ and $\bar{a}$ as follows. Let the singular value decomposition of $\bar{h}(I - \bar{a})^{-1}$ be UIV^T, then $\hat{g}_i$ are the elements of $U^T g$ and γ_i are the diagonal entries of Γ, where g is the observed image and T denotes the complex conjugate transpose operator. The expected value of the GCV criterion attains a global minimum at the true values of the ARMA model parameters [38]. In practice, due to the huge sizes of $\bar{h}$ and $\bar{a}$ (each has the size of $N^2 \times N^2$), Eq. (3.11) is often implemented in the frequency domain as:

$$V(\theta) = \frac{\frac{1}{N^2} \sum_{p=0}^{N-1} \sum_{q=0}^{N-1} \left(1 - \frac{(conj(H) * H)_{p,q}}{(conj(H) * H)_{p,q} + \alpha (conj(B) * B)_{p,q}} \right)^2 |G_{p,q}|^2}{\left[1 - \frac{1}{N^2} \sum_{p=0}^{N-1} \sum_{q=0}^{N-1} \frac{(H'H)_{p,q}}{(conj(H) * H)_{p,q} + \alpha (conj(B) * B)_{p,q}} \right]^2} \tag{3.12}$$

where H, A, and G are 2-D DFT of h, a and g, respectively. *conj*(H) and *conj*(B) are the complex conjugate of H and B, with $B=I-A$. Here we have assumed that the matrices A and H are circulant. It implies that the image has been blurred by a circular convolution operation. This assumption has been used in many traditional image restoration algorithms [38, 122] to reduce the computational complexity. As long as the image boundaries are properly preprocessed to remove false boundary discontinuities, the circular convolution assumption has a negligible effect on the identification results. Besides, it is further assumed in many traditional approaches that most images can be represented adequately as a process whose auto-correlation function consists of a separable exponentially decaying sequence [38, 122]. It results in a much simplified AR model that can be described by only two parameters as follows: $a_{10}=\rho_v$, $a_{01}=\rho_h$, and $a_{11}=-\rho_v\rho_h$; where a_{ij} represents the AR formation model parameter for the original image. Moreover, the blurring process is assumed to be energy conservative, that is $\sum_{(k,l)\in R_h} h_{kl} = 1$; where R_h is the support of the blur function. The energy conservation assumption is valid due to the fact that a fixed number of photons are detected regardless of whether blurring occurs.

3.4 Efficient Blind Image Restoration Using DPRT

3.4.1 Converting 2-D GCV to 1-D GCVs using DPRT

The GCV criterion is clearly nonlinear in the parameters of interest and cannot be minimized analytically. Therefore, numerical techniques are used to determine the parameter set. It has been shown in [38] that the numerical search performs particularly well for single parameter PSF. However, in case of blurring due to multiple parameter PSF, it is often required to carry out numerical optimization procedure for more than a few hundreds of iterations before the procedure converges. It requires a long computation time since, for

each iteration, the GCV score for a 2-D signal is to be determined. In addition, the huge volume of memory transfers due to the buffering of 2-D signals also requires a significant amount of computation time.

Since DPRT is able to convert a 2-D circular convolution problem into 1-D ones, it is intuitively to consider using DPRT for blind image restoration. The idea is very simple. Recall Eqs.(3.9a) and (3.9b) and rewrite them using the circular convolution operator,

$$f = \tilde{a} \otimes_2 f + u \tag{3.13}$$

$$g = \tilde{h} \otimes_2 f + n. \tag{3.14}$$

Note that the supports of PSF and the AR model are often much smaller than the image. The function $\tilde{h}$ and $\tilde{a}$ in Eqs. (3.13) and (3.14) refer to the original h and a padded with the appropriate number of zeros to enable them to have the same size as the image support. Assume that the DPRT of g, f, $\tilde{h}$, $\tilde{a}$, u and n are $\{g_m^c, g_s^b\}$, $\{f_m^c, f_s^b\}$, $\{\tilde{h}_m^c, \tilde{h}_s^b\}$, $\{\tilde{a}_m^c, \tilde{a}_s^b\}$, $\{u_m^c, u_s^b\}$, and $\{n_m^c, n_s^b\}$, respectively, where m = 0,..., 2^n-1; s = 0,..., 2^{n-1}-1 and the sizes of all functions are $N \times N$, where $N = 2^n$; $n \in \mathbf{Z}$. Then, from the circular convolution property of DPRT, we have

$$f_m^c = f_m^c \otimes \tilde{a}_m^c + u_m^c \qquad g_m^c = f_m^c \otimes \tilde{h}_m^c + n_m^c$$

$$f_s^b = f_s^b \otimes \tilde{a}_s^b + u_s^b \qquad g_s^b = f_s^b \otimes \tilde{h}_s^b + n_s^b. \tag{3.15}$$

Eq. (3.15) shows that the original 2-D blind image restoration problem is converted into $3N/2$ 1-D blind image restoration problems. 1-D GCV algorithm can be applied to estimate $\{\tilde{h}_m^c, \tilde{h}_s^b\}$ and $\{\tilde{a}_m^c, \tilde{a}_s^b\}$ from $\{g_m^c, g_s^b\}$. The 1-D GCV score for the mth projection becomes:

$$V(\theta_m) = \frac{\frac{1}{N}\sum_{p=0}^{N-1}\left(1 - \frac{(conj(\tilde{H}_m^c) * \tilde{H}_m^c)_p}{(conj(\tilde{H}_m^c) * \tilde{H}_m^c)_p + \alpha_m (conj(\tilde{B}_m^c) * \tilde{B}_m^c)_p}\right)^2 \left|G_m^c(p)\right|^2}{\left[1 - \frac{1}{N}\sum_{p=0}^{N-1} \frac{(conj(\tilde{H}_m^c) * \tilde{H}_m^c)_p}{(conj(\tilde{H}_m^c) * \tilde{H}_m^c)_p + \alpha_m (conj(\tilde{B}_m^c) * \tilde{B}_m^c)_p}\right]^2} \tag{3.16}$$

where $\tilde{H}_m^c$ and G_m^c are the DFT of $\tilde{h}_m^c$ and g_m^c, respectively. $\tilde{B}_m^c$ is the DFT of $\tilde{b}_m^c$, with $\tilde{b}_m^c = I - \tilde{a}_m^c$. The parameter set θ_m in this case becomes $\{\alpha_m, \tilde{h}_m^c, \tilde{a}_m^c\}$. The 1-D GCV score for the sth projections is similar to Eq. (3.16) with all m replaced by s. Let us summarize the whole procedure as follows:

1. Transform the observed image g into the DPRT domain to become $\{g_m^c, g_s^b\}$.
2. Apply 1-D GCV algorithm to estimate $\{\tilde{h}_m^c, \tilde{h}_s^b\}$ and $\{\tilde{a}_m^c, \tilde{a}_s^b\}$ from $\{g_m^c, g_s^b\}$.
3. Reconstruct $\tilde{h}$ and $\tilde{a}$ based on $\{\tilde{h}_m^c, \tilde{h}_s^b\}$, $\{\tilde{a}_m^c, \tilde{a}_s^b\}$ using the inverse DPRT.
4. Obtain the restored image using any computationally inexpensive restoration algorithm based on the estimated $\tilde{h}$ and $\tilde{a}$.

3.4.2 Proposed DPRT blind image restoration algorithm

Unfortunately, the intuitive idea suggested in Section 3.4.1 does not work properly in general. Due to the non-orthogonal property of DPRT, the $3N/2$ 1-D GCV restorations have the total computational complexity not so much less than the original 2-D problem. Furthermore, as it is mentioned above, the DPRT of a compact-supported function may not be compact-supported. It requires to estimate more parameters than the original 2-D approach. Let us use an example to illustrate the problem. Assume that g has the size of 256×256. Assume also that h and a have the size of 5×5 and 2×2, respectively. Hence, only 30 parameters are required to be estimated in the parameter set originally. In Eq. (3.15), the support sizes of $\{\tilde{h}_m^c, \tilde{h}_s^b\}$ and $\{\tilde{a}_m^c, \tilde{a}_s^b\}$ can be up to 256, i.e. the size of the image support. It implies that for each m and s, up to 256 parameters are required to be estimated. It is simply unachievable with the GCV algorithm. Thus, it is known that direct blind image restoration using DPRT is not feasible. Consequently, we modify the algorithm such that only a few

essentially compact-supported projections are used for the restoration. In this way, the computational complexity is greatly reduced.

It is interesting to note that, depending on the size of h and $\tilde{h}$, some of the projections of $\tilde{h}$ are compact-supported. It is shown in the following lemma:

Lemma 3.1: Given that a function $\tilde{h}$ with size $N\times N$ is formed by padding appropriate number of zeros to another function h with size $P\times P$, where $N = 2^n$; $n\in\mathbf{Z}$ and P is a prime integer. If the DPRT of $\tilde{h}$ is $\{\tilde{h}_m^c(d), \tilde{h}_s^b(d)\}$, where $m, d = 0,..., N\text{-}1$; $s = 0,..., N/2\text{-}1$, then

$$\begin{aligned}
\tilde{h}_m^c(d) = 0, \quad & \text{if} \quad \left\{ P \le d \le N-1-m(P-1) \ \text{ and } \ m < \frac{N-P}{P-1} \right\} \\
& \text{or if} \quad \left\{ P+(N-m)(P-1) \le d \le N-1 \ \text{ and } \ m > N - \frac{N-P}{P-1} \right\} \\
\tilde{h}_s^b(d) = 0, \quad & \text{if} \quad \left\{ P \le d \le N-1-2s(P-1) \ \text{ and } \ s < \frac{N-P}{2(P-1)} \right\} \\
& \text{or if} \quad \left\{ P+(N-2s)(P-1) \le d \le N-1 \ \text{ and } \ s > \frac{N}{2} - \frac{N-P}{2(P-1)} \right\}.
\end{aligned} \tag{3.17}$$

The proof of Lemma 3.1 is shown in Appendix A. In fact, Lemma 3.1 only describes a special case that N is a power of 2 and P is a prime number. A more general lemma can be derived to deal with different combinations of N and P. However, the case for Eq. (3.17) is typical for many image restoration problems. Lemma 3.1 shows that when m and s are very small or very big numbers, there will be a consecutive sequence of zeros in the projections. This implies that the first and the last few projections of the total $3N/2$ projections are essentially compact-supported. From Eq. (3.17), we can also derive the support sizes of

some of the essentially compact-supported projections. For instance, if $S(.)$ refers to the support size of a projection, we know that

$$\begin{aligned} S(\tilde{h}_m^c) &= P + m(P-1) && \text{for } m < \frac{N-1}{P-1} \\ S(\tilde{h}_m^c) &= P + (N-m)(P-1) && \text{for } m > N - \frac{N-1}{P-1} \\ S(\tilde{h}_s^c) &= P + 2s(P-1) && \text{for } s < \frac{N-1}{2(P-1)} \\ S(\tilde{h}_s^c) &= P + 2(\frac{N}{2} - s)(P-1) && \text{for } s > \frac{N}{2} - \frac{N-1}{2(P-1)}. \end{aligned} \tag{3.18}$$

Fig. 3.2 shows an example to illustrate Lemma 3.1. In Fig. 3.2(a), a 3×3 2-D function h is padded with zeros to form an 8×8 function $\tilde{h}$. The DPRT of the zero-padded function is shown in Fig. 3.2(b). It is seen that the first two and the last columns (projections) of $\tilde{h}_m^c$ and the first column of $\tilde{h}_s^b$ are essentially compact-supported.

$$\begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \xrightarrow{\textbf{\textit{DPRT}}} \begin{pmatrix} 0 & 2 & 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 2 & 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 2 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 & 2 \\ 3 & 1 & 1 & 2 & 2 & 1 & 2 & 3 \\ 3 & 2 & 1 & 1 & 2 & 1 & 1 & 2 \\ 3 & 3 & 2 & 1 & 2 & 2 & 1 & 1 \end{pmatrix} \quad \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 2 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 2 \\ 0 & 0 & 0 & 1 \\ 3 & 1 & 2 & 2 \\ 3 & 1 & 2 & 1 \\ 3 & 2 & 2 & 1 \end{pmatrix}$$

$\tilde{h}_m^c(d)$ $\tilde{h}_s^b(d)$

(a) Original 2-D function $\tilde{h}$ (b) The DPRT of $\tilde{h}$

Figure 3.2: An example to illustrate Lemma 3.1.

In the previous example, it is shown that some of the projections are essentially compact-supported. When applying the 1-D GCV algorithm to these essentially compact-supported projections, the number of parameters to be estimated is much less than the other projections. We now further show that only some of these essentially compact-supported projections are useful for reconstructing the original compact-supported PSF h. More specifically, we show in Lemma 3.2 that, if we acquire P essentially compact-supported projections from $\widetilde{h}_m^c$ and the first projection from $\widetilde{h}_s^b$, we can obtain $\{h_m^c, h_0^b\}$ (i.e. the DPRT of h) by some simple additions on these projections. Consequently, we can reconstruct h from $\{h_m^c, h_0^b\}$ based on the inverse DPRT as shown in Eq. (3.6).

Lemma 3.2: Given that a function $\widetilde{h}$ with size $N\times N$ is formed by padding appropriate number of zeros to another function h with size $P\times P$, where $N = 2^n$; $n \in \mathbf{Z}$ and P is a prime integer. Assume that the DPRT of $\widetilde{h}$ is $\{\widetilde{h}_m^c(d), \widetilde{h}_s^b(d)\}$, where m,d=0,...,N-1; s = 0,...,N/2-1, and the DPRT of h is $\{h_m^c(d), h_0^b(d)\}$, where m, d = 0,..., P-1. Assume also that $N > (P^2+1)$, then

(i) $h_0^b(d) = \widetilde{h}_0^b(d)$ for d = 0,...,P-1.

(ii) $h_m^c(d) = \sum_{n=0}^{m-\lfloor m/P \rfloor} \widetilde{h}_m^c(<d+P(N-n)>_N)$ for $0 \le m < P/2$ and $d = 0,...,P-1$.

(iii) $h_{P-1-m}^c(d) = \sum_{n=0}^{m-\lfloor m/P \rfloor} \widetilde{h}_{N-1-m}^c(<d+Pn>_N)$ for $0 \le m < P/2$ and $d = 0,...,P-1$.

(3.19)

The proof of Lemma 3.2 is shown in Appendix B. Lemma 3.2 shows that by appropriately adding the data of the essentially compact-supported projections obtained from $\{\tilde{h}_m^c, \tilde{h}_s^b\}$, we can reconstruct $\{h_m^c, h_0^b\}$. More importantly, only P+1 projections are required for the reconstruction of $\{h_m^c, h_0^b\}$ as compared to the original $3N/2$ projections. The computational complexity is greatly reduced.

Eq. (3.19) allows us to evaluate $\{h_m^c, h_0^b\}$ from $\{\tilde{h}_m^c, \tilde{h}^b\}$ and then h can be obtained from $\{h_m^c, h_0^b\}$ using the inverse DPRT. Besides estimating the PSF, the estimation of the image model parameters a is also important. The procedure for evaluating a can be much simplified due to the assumption we made. Recall that images are assumed to be represented adequately by an AR model a that can be described by only two parameters as follows: $a_{10} = \rho_v$, $a_{01} = \rho_h$, and $a_{11} = -\rho_v\rho_h$; where a_{ij} represents the parameters of the AR formation model for the original image. With this simple structure, all parameters can be easily obtained from the DPRT projections of $\tilde{a}$. More specifically, it is easy to show that

$$\tilde{a}_0^c(0) = \rho_v, \tilde{a}_0^b(0) = \rho_h. \tag{3.20}$$

Hence once $\tilde{a}_0^c$ and $\tilde{a}_0^b$ are obtained from the 1-D GCV algorithm, ρ_v and ρ_h can also be obtained. When both h and a are available, the only thing missing is the regularization parameter α. It can be easily obtained by a direct search and using the 2-D GCV score as the stopping criterion.

Let us summarize the proposed DPRT blind image restoration algorithm as follows. Recall Eqs. (3.13) and (3.14) for g and h with sizes equal to $N \times N$ and $P \times P$, respectively, where $N = 2^n$; $n \in \mathbf{Z}$ and P is a prime integer. Then, with proper boundaries adjustment of g and f such that the circular convolutions in Eqs. (3.13) and (3.14) are equivalent to the linear convolutions in Eqs. (3.9a) and (3.10), the proposed DPRT blind image restoration algorithm is as follows:

1. Perform DPRT on g to obtain $\{g_m^c, g_s^b\}$ which is equal to the circular convolutions of $\{f_m^c, f_s^b\}$ and $\{\tilde{h}_m^c, \tilde{h}_s^b\}$, respectively, as in Eq. (3.15).
2. Select g_0^b and P other projections from g_m^c of which the corresponding projections of $\tilde{h}_m^c$ are essentially compact-supported. More specifically, we select the first $\lceil P/2 \rceil$ and the last $\lfloor P/2 \rfloor$ projections of g_m^c and projection g_0^b.
3. 1-D GCV algorithm is then applied to identify $\{\tilde{h}_m^c, \tilde{h}_0^b\}$ and $\{\tilde{a}_m^c, \tilde{a}_0^b\}$ for the selected projections.
4. Reconstruct $\{h_m^c, h_0^b\}$ from $\{\tilde{h}_m^c, \tilde{h}_0^b\}$ using Eq. (3.19).
5. Obtain a from $\{\tilde{a}_0^c, \tilde{a}_0^b\}$ using Eq. (3.20).
6. Obtain h from $\{h_m^c, h_0^b\}$ using the inverse DPRT algorithm as stated in Eq. (3.6).
7. Based on the estimated h and a, obtain α by direct searching and using the 2-D GCV score as the stopping criterion.
8. Restore the image using the estimated h, a, and α.

A block diagram is given below to illustrate the procedure:

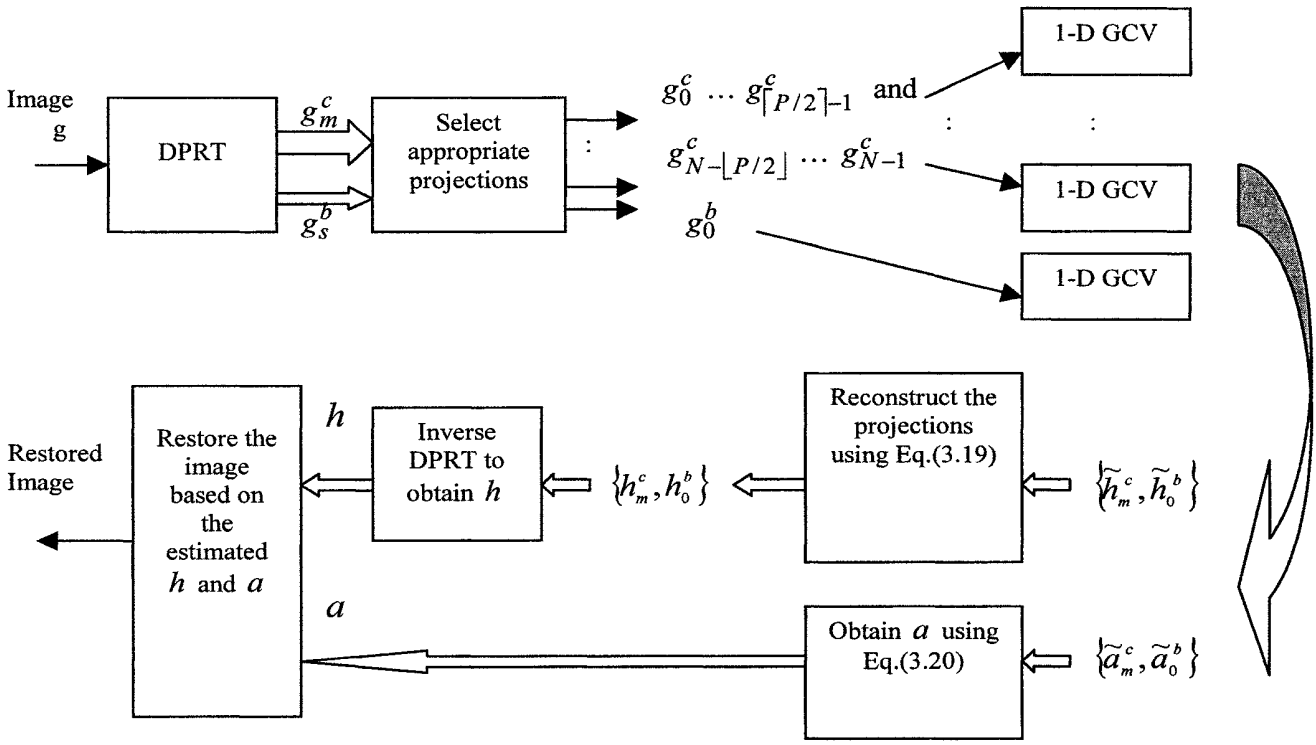


Figure 3.3: A block diagram that describes the flow of the proposed algorithm.

3.4.3 Implementation issues

Similar to the other traditional approaches [38, 122], we only consider symmetric blur models. It is known that many real-life blurs, such as the linear motion, out-of-focus, and atmospheric turbulence blurs, are symmetric in nature. This assumption is not restrictive but greatly reduces the search space and allows the optimization procedure to converge more easily.

When the PSF is symmetric, we can easily show that its DPRT projections are also symmetric up to a rotation. Furthermore, some projections differ from the other only by a rotation. All these imply that the number of parameters to be estimated can be greatly

reduced in the DPRT domain. Let us formally summarize these properties of DPRT as follows:

Lemma 3.3: Given that a function $\tilde{h}$ with size $N\times N$ is formed by padding appropriate number of zeros to another function h with size $P\times P$, where $N = 2^n$; $n \in \mathbf{Z}$ and P is a prime integer. Assume that the DPRT of $\tilde{h}$ is $\{\tilde{h}_m^c(d), \tilde{h}_s^b(d)\}$, where m,d=0,...,N-1; s = 0,...,N/2-1, and the DPRT of h is $\{h_m^c(d), h_0^b(d)\}$, where m, d = 0,..., P-1. If h is symmetric such that

$$h(x,y) = h(<P-x-1>_P, <P-y-1>_P)$$

where x, y = 0,..., P-1, then

1. $\{\tilde{h}_m^c(d), \tilde{h}^b(d)\}$ is also symmetric up to a rotation.
2. $\tilde{h}_0^c(d) = \tilde{h}_0^b(d)$.
3. $abs\left(\text{FFT}\{\tilde{h}_m^c(d), \tilde{h}_s^b(d)\}\right) = abs\left(\text{FFT}\{\tilde{h}_{N-m}^c(d), \tilde{h}_{N/2-s}^b(d)\}\right)$,

   where m = 1,..., N-1; s = 1,...,N/2-1.

The proof of Lemma 3.3 is shown in Appendix C. Item 1 of Lemma 3.3 implies that, for every projection in this case, only half of the coefficients are required to be estimated. The other half is the same as the first half. Item 2 of Lemma 3.3 implies that the coefficients of $\tilde{h}_0^c$ are the same as $\tilde{h}_0^b$. Hence, only one set of them needs to be estimated. Item 3 of Lemma 3.3 implies that for the P projections required to be estimated as indicated in Lemma 2, (P-1)/2 of them need not be estimated if h is symmetric. It is because they have the same coefficients as the other (P-1)/2 projections.

Let us use an example to further illustrate the actual numbers of coefficients need to be estimated for various projections. Let g has the size of 256×256 and h has the size of

5×5. Based on Lemma 3.2, we know that only $g_0^c, g_1^c, g_2^c, g_{254}^c, g_{255}^c$ and g_0^b are required to estimate *h*. From Lemma 3.3, we know that $\tilde{h}_0^c$ and $\tilde{h}_0^b$ have the same coefficients. Hence we only need to perform 1-D GCV algorithm on g_0^c to obtain both $\tilde{h}_0$ and $\tilde{h}_0^b$. Furthermore, we know from Lemma 3.3 that $\{\tilde{h}_1^c, \tilde{h}_2^c\}$ and $\{\tilde{h}_{254}^c, \tilde{h}_{255}^c\}$ have the same set of coefficients. Again we only need to perform 1-D GCV algorithm on $\{g_1^c, g_2^c\}$ to obtain both $\{\tilde{h}_1^c, \tilde{h}_2^c\}$ and $\{\tilde{h}_{254}^c, \tilde{h}_{255}^c\}$. Finally, when estimating $\{\tilde{h}_0^c, \tilde{h}_1^c, \tilde{h}_2^c\}$ from the projections $\{g_0^c, g_1^c, g_2^c\}$, Lemma 3.3 indicates that half of the coefficients of $\{\tilde{h}_0^c, \tilde{h}_1^c, \tilde{h}_2^c\}$ are the same as the other due to the symmetric property. From Eq. (3.18), we know that the supports of $\{\tilde{h}_0^c, \tilde{h}_1^c, \tilde{h}_2^c\}$ are 5, 9, and 13, respectively. Due to the symmetry property, only 3, 5, and 7 of them, respectively, need to be estimated.

3.5 Simulation Results

To verify the improvement achieved by the proposed approach, the traditional 2-D GCV blind image restoration algorithm and the proposed DPRT based algorithm are implemented and compared. Both algorithms are implemented using Matlab and the command *fmincon* is used in both cases for numerical search. To improve the convergence rate of both approaches, a nonlinear constraint is imposed to the PSF during estimation. The nonlinear constraint is basically a band-reject filter to minimize the high frequency component of the PSF and we find that this constraint is particularly important to the 2-D GCV algorithm. For both approaches, the following 5×5 PSF is used:

$$\begin{bmatrix} 0.0100 & 0.0248 & 0.0287 & 0.0248 & 0.0100 \\ 0.0248 & 0.0377 & 0.0812 & 0.0377 & 0.0248 \\ 0.0287 & 0.0812 & 0.1714 & 0.0812 & 0.0287 \\ 0.0248 & 0.0377 & 0.0812 & 0.0377 & 0.0248 \\ 0.0100 & 0.0248 & 0.0287 & 0.0248 & 0.0100 \end{bmatrix}$$

The initial conditions for the 2-D GCV approach are:

$$\{\alpha = 0.0001;\ \rho_v = 0.8;\ \rho_h = 0.8;\ h_{00} = 1;\ h_{ij} = 0;\ \text{for}\ (i,j) \neq (0,0)\}$$

The initial conditions for the proposed DPRT based algorithm are:

$$\{\alpha_m = 0.0001;\ \rho_v = 0.8;\ \rho_h = 0.8;\ \tilde{h}_m^c(0) = 1;\ \tilde{h}_m^c(d) = 0;\ \text{for}\ d \neq 0\}$$

$$\{\alpha_s = 0.0001;\ \rho_v = 0.8;\ \rho_h = 0.8;\ \tilde{h}_0^b(0) = 1;\ \tilde{h}_0^b(d) = 0;\ \text{for}\ d \neq 0\}$$

They are selected rather arbitrarily. Different standard testing images, such as, Lenna and Pepper, etc. are used and blurred by the PSFs. Additive white Gaussian noise is then added to the blurred images in two different noise levels: BSNR = 30dB and 40dB according to Eq. (2.7).

For each image, 20 experiments are performed for each approach at each noise level. The results in terms of accuracy and complexity in all experiments are recorded and averaged. The accuracy is measured by the signal-to-error ratio (SER) between the restored image and the original image. The accuracy in estimating the PSF is also measured in terms of SER. The complexity of the algorithms is measured by the required number of floating point operations (flops) as reported by the *flops* command of Matlab. Since the major computation is in the numerical search, the number of flops in each iteration and the number of iterations for both approaches are recorded and compared. Table 3.1 shows a comparison of the computational complexity of the 2-D GCV algorithm and the proposed DPRT-based algorithm.

	2-D GCV			Proposed DPRT approach		
Noise Level (BSNR)	No. of iterations (I)	No of. Flops per iter. (F)	I×F	No. of iterations (I)	No of. flops per iter. (F)	I×F + DPRT
40dB	361	5,554,950	2.01×10^9	1,515	40,759	6.25×10^7
30dB	318	5,554,950	1.76×10^9	1,648	40,759	6.8×10^7

Table 3.1: Comparison in terms of arithmetic operations required.

Note that the result on "number of iterations" of the DPRT approach is the sum of all iterations required for the estimation of the parameter set in all projections. Note also that the floating point operations required for the implementation of DPRT itself are also included and added to the results of the proposed DPRT-based approach. Table 3.1 shows that, in both noise levels, the total number of operations required for the proposed approach is only about 4% of the 2-D GCV algorithm. This result is foreseeable because the major operation done in each iteration of the GCV algorithm is an FFT. By converting the 2-D GCV algorithm to become some 1-D ones, the $N\times N$-point 2-D FFT required in each iteration is also converted into some N-point 1-D FFTs. For an image with size 256×256, a 256-point 1-D FFT requires only 0.2% of floating point operations as compared with a 256×256-point 2-D FFT. Therefore, although a few more 1-D GCV algorithms need to be implemented in the proposed approach, the overall complexity is still only 4% of the 2-D case. Indeed the actual computation time required by the 2-D approach can be even longer than the proposed DPRT algorithm due to the massive 2-D data transfer, of which the burden introduced has not been reflected in the number of arithmetic operations reported. This kind of data movement is particularly slow in Matlab if it is not carefully handled. Table 3.2 illustrates the accuracy of both approaches.

	2-D GCV		Proposed DPRT approach	
Noise level (BSNR)	Restored image accuracy (SER)	Estimated PSF accuracy (SER)	Restored image accuracy (SER)	Estimated PSF accuracy (SER)
40dB	25.82 dB	14.16 dB	26.78 dB	14.12 dB
30dB	23.10 dB	12.68 dB	23.42 dB	13.84 dB

Table 3.2: Comparison in terms of accuracy achieved.

The results in Table 3.2 show that the accuracy of the proposed DPRT approach is not inferior to the traditional 2-D approach, particularly in high noise level. Both approaches have the problem of jumping into local minimum during numerical search. The problem happened more often when noise level is lowered to BSNR = 40dB. For the proposed approach, this problem happens about 5 times in each 20 experiments. However, the problem can often be solved by restarting the algorithm with the estimated parameters serving as the initial guess plus a random turbulence. The results for model parameter estimations are not shown here since the estimation results varied from one experiment to another. This happens to both approaches since the image model, anyway, is only an approximation to the actual one. Figs. 3.4 and 3.5 show the actual restored images given by both approaches. They can be considered as a subjective measure for the accuracy of the algorithms. However, it is seen that no observable difference can be found from the restored images obtained by the two approaches. Actually, the level of distortion introduced is typical and is the same as in many other works reported in the literature. For the case of lower SNR and/or more severe distortion, the performance of the proposed algorithm may be degraded in a sense that more iterations may be required or the estimation does not converge. However, in most of the cases, the proposed approach performs better than the traditional ones at the same distortion level. As to the order of blurring function, it is

assumed that, as in other research works, such parameter is known before the estimation of the blurring function. When an incorrect order of blurring function is applied, more iterations may be required or the estimation does not converge.

It should be noted that other fast multidimensional convolution algorithms, such as the polynomial transform [123], can also be applied to convert the 2-D operation to several 1-D operations for 2-D GCV instead of using DPRT. However, as mentioned before, the transformed data of a compact-supported function may not be compact-supported after applying the transform. In this case, even though we have converted the 2-D operations into some 1-D ones, the resulted computational complexity would not be much lesser than that of the original 2-D problem as the size of the resulted parameter sets required for estimation is not much reduced. This is the reason why we try to minimize the parameter set of the DPRT of a compact-supported function required for restoration.

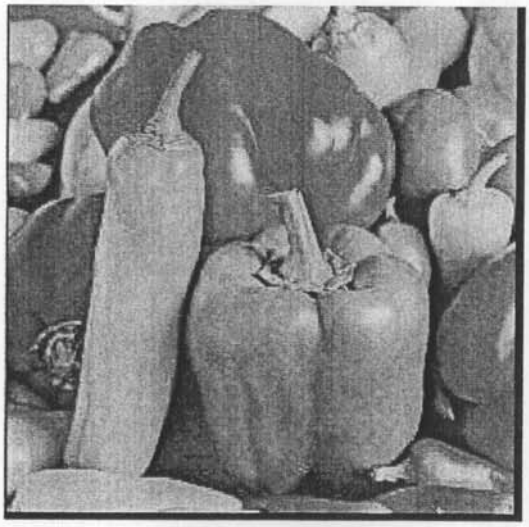
(a)

(b)

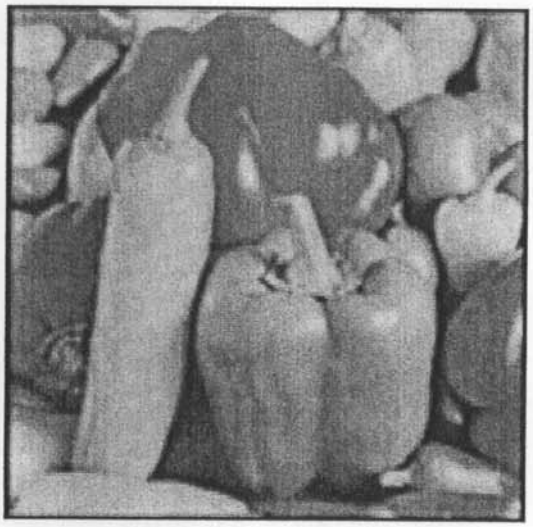
(c)

(d)

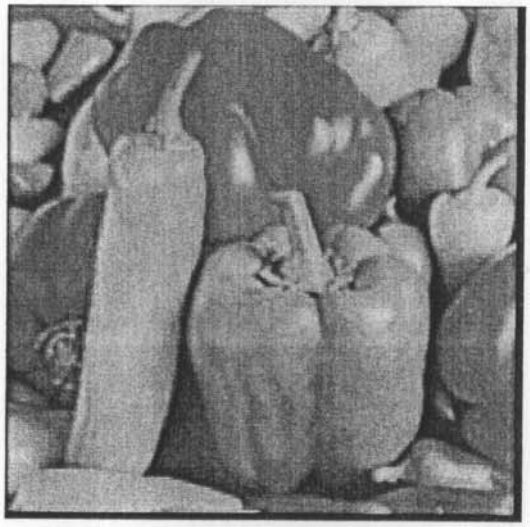
Figure 3.4: (a) Original image (256×256); (b) blurred and noisy image with BSNR = 30dB; restored image using 2-D GCV with SER = 22.01dB and (d) restored image using the proposed DPRT with SER = 22.66dB.



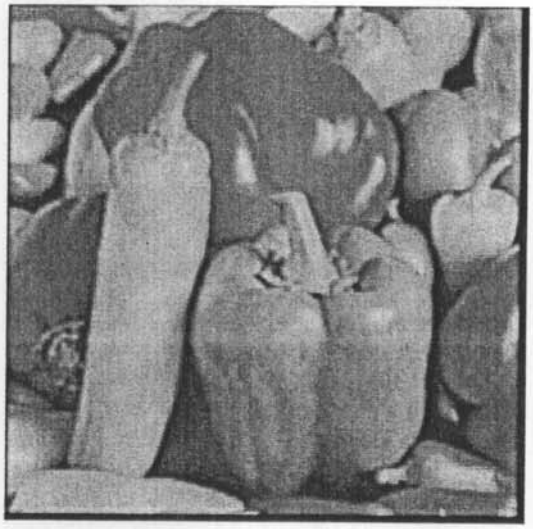
(a)



(b)



(c)



(d)

Figure 3.5: (a) Original image (256×256); (b) blurred and noisy image with BSNR = 30dB; (c) restored image using 2-D GCV with SER = 23.62dB and (d) restored image using the proposed DPRT with SER = 23.57dB.

3.6 Summary

In this chapter, efficient algorithm using DPRT is proposed for blind image restoration. By using DPRT, the original 2-D blind image restoration problem is converted to some 1-D ones hence greatly reduces the memory size and computation time. Experimental results show that the proposed DPRT-based approach consistently uses less arithmetic operations than the 2-D GCV approach. The overall saving can be as much as 96%. Apart from the saving in computation time, the accuracy of the proposed approach is not inferior to the traditional one in both objective and subjective measures. While the proposed approach is only a fast algorithm to the original one, it shares the same limitations. For example, the GCV algorithm can only estimate the magnitude of the PSF but not the phase information. Besides, as most of the traditional approaches, the support of the PSF is assumed to be known in the proposed algorithm. This assumption may not be valid in some practical situations. It means that a priori process may be required to estimate the support of the PSF. Some results from the work on ARMA model order estimation [124] can be used to deal with this problem. Further work is being carried out to tackle these two problems.

Appendix A

Proof of Lemma 3.1

Given that a function $\tilde{h}$ with size $N\times N$ is formed by padding appropriate number of zeros to another function h with size $P\times P$, where $N = 2^n$; $n \in \mathbf{Z}$ and P is a prime integer. Given also that the DPRT of $\tilde{h}$ is $\{\tilde{h}_m^c(d), \tilde{h}_0^b(d)\}$ such that

$$\tilde{h}_m^c(d) = \sum_{x=0}^{N-1} \tilde{h}(x, < d + mx >_N) \tag{A.3.1}$$

$$\tilde{h}_s^b(d) = \sum_{y=0}^{N-1} \tilde{h}(< d + 2sy >_N, y) \tag{A.3.2}$$

where $m, d = 0,..., N-1; s = 0,..., N/2-1$. From Eq. (A.3.1), we know that

$$\tilde{h}_m^c(d) = \sum_{x=0}^{N-1} \tilde{h}(x, < d + mx >_N) = \sum_{x=0}^{P-1} \tilde{h}(x, < d + mx >_N) + \sum_{x=P}^{N-1} \tilde{h}(x, < d + mx >_N)$$

Since

$$\tilde{h}(x, y) = 0 \qquad if \quad x, y > P-1$$

Hence if

$$\sum_{x=0}^{P-1} \tilde{h}(x, < d + mx >_N) = 0 \quad \Rightarrow \quad \tilde{h}_m^c(d) = 0$$

Or equivalently if $< d + mx >_N \geq P \quad \forall\, x \leq P-1 \quad \Rightarrow \quad \tilde{h}_m^c(d) = 0$

Let us consider for all m such that $m(P-1) < N-P$. Hence $N-1-m(P-1) \geq P$. Consider for all d such that $P \leq d \leq N-1-m(P-1)$. Since $0 \leq x \leq P-1$,

$$P \leq d + mx \leq N-1 \qquad \text{if} \quad P \leq d \leq N-1-m(P-1) \tag{A.3.3}$$

That is,

$$< d + mx >_N = d + mx \geq P \qquad \text{if} \quad P \leq d \leq N-1-m(P-1) \ \text{ and } \ m < \frac{N-P}{P-1}$$

Hence $\tilde{h}_m^c = 0 \qquad \text{if} \quad P \leq d \leq N-1-m(P-1) \ \text{ and } \ m < \dfrac{N-P}{P-1}$ (Q.E.D.)

Let us consider for all m such that $m > N - \frac{N-P}{P-1}$. Hence $P + (N-m)(P-1) \le N-1$.

Consider for all d such that $P + (N-m)(P-1) \le d \le N-1$.

$$P + (N-m)(P-1) + mx \le d + mx \qquad \text{if} \quad P + (N-m)(P-1) \le d \le N-1$$

Let $k = N - m$,

$$P + k(P-1) + (N-k)x \le d + (N-k)x \qquad \text{if} \quad P + k(P-1) \le d \le N-1$$

Since $x \le P-1$,

$$P + Nx \le d + (N-k)x \qquad \text{if} \quad P + k(P-1) \le d \le N-1$$

Hence

$$P \le d + mx - Nx \qquad \text{if} \quad P + (N-m)(P-1) \le d \le N-1$$

Since $0 \le x$ and $d \le N-1$,

$$\begin{aligned} d + mx - Nx &= d - (N-m)x \\ &\le N-1 \end{aligned}$$

Hence

$$P \le d + mx - Nx \le N-1 \qquad \text{if} \quad P + (N-m)(P-1) \le d \le N-1 \quad \text{and} \quad m > N - \frac{N-P}{P-1}$$

That is,

$$< d + mx >_N \ge P \qquad \text{if} \quad P + (N-m)(P-1) \le d \le N-1 \quad \text{and} \quad m > N - \frac{N-P}{P-1}$$

Hence $\widetilde{h}_m^c = 0 \qquad \text{if} \quad P + (N-m)(P-1) \le d \le N-1 \quad \text{and} \quad m > N - \frac{N-P}{P-1}$

(Q.E.D.)

The proof for $\widetilde{h}_s^b$ is similar to that described above.

Appendix B

Proof of Lemma 3.2

Given that a function $\widetilde{h}$ with size $N \times N$ is formed by padding appropriate number of zeros to another function h with size $P \times P$, where $N = 2^n$; $n \in \mathbf{Z}$ and P is a prime integer. It is known that $N > P^2 + 1$, and the DPRT of $\widetilde{h}$ is $\{\widetilde{h}_m^c(d), \widetilde{h}_s^b(d)\}$ such that

$$\widetilde{h}_m^c(d) = \sum_{x=0}^{N-1} \widetilde{h}(x, < d + mx >_N) \tag{B.3.1}$$

$$\widetilde{h}_s^b(d) = \sum_{y=0}^{N-1} \widetilde{h}(< d + 2sy >_N, y) \tag{B.3.2}$$

where $m, d = 0,\ldots, N\text{-}1$; $s = 0,\ldots, N/2\text{-}1$. The DPRT of h is $\{h_m^c(d), h_0^b(d)\}$ such that

$$h_m^c(d) = \sum_{x=0}^{P-1} h(x, < d + mx >_P) \tag{B.3.3}$$

$$h_0^b(d) = \sum_{y=0}^{P-1} h(d, y) \tag{B.3.4}$$

where $m, d = 0,\ldots, P\text{-}1$. Since

$$\widetilde{h}(x, y) = 0 \qquad \text{if} \quad x, y > P - 1 \tag{B.3.5}$$

we immediately have the following result:

$$h_0^b(d) = \widetilde{h}_0^b(d) \qquad \text{for} \quad d=0,\ldots,P\text{-}1$$

Hence Lemma 3.2 case (i) is proved. (Q.E.D.)

The relationship between f_m^c and $f_m^{c'}$ can be proved as follows:

Since

$$\tilde{h}_m^c(d) = \sum_{x=0}^{N-1} \tilde{h}(x, < d + mx >_N)$$
$$= \sum_{x=0}^{P-1} \tilde{h}(x, < d + mx >_N) + \sum_{x=P}^{N-1} \tilde{h}(x, < d + mx >_N)$$

from Eq. (B.3.5), we know that the second term is equal to 0. Hence

$$\tilde{h}_m^c(d) = \sum_{x=0}^{P-1} \tilde{h}(x, < d + mx >_N) \tag{B.3.6}$$

For m = 0,..., (P-1)/2,

$$\sum_{n=0}^{m-\lfloor m/P \rfloor} \tilde{h}_m^c(< d + (N-n)P >_N) = \sum_{n=0}^{m-\lfloor m/P \rfloor} \sum_{x=0}^{P-1} \tilde{h}(x, < d + mx - nP >_N)$$
$$= \sum_{x=0}^{P-1} \sum_{n=0}^{m-\lfloor m/P \rfloor} \tilde{h}(x, < d + mx - nP >_N) \tag{B.3.7}$$

Eq. (B.3.7) shows that if

$$\sum_{n=0}^{m-\lfloor m/P \rfloor} \tilde{h}(x, < d + mx - nP >_N) = h(x, < d + mx >_P) \tag{B.3.8}$$

then the proof of Lemma 3.2 case(ii) is completed. Indeed, the maximum value of the term d+mx can also be determined to be:

$$d + mx \le P - 1 + (P-1)^2/2 \le (P^2 - 1)/2 < N$$

Hence, since both d, x and m are positive integers, the term $h(x, < d + mx >_P)$ can be rewritten as follows:

$$h(x, < d + mx >_P) = h(x, < d + mx - iP >_N) \tag{B.3.9}$$

for an integer i. From the maximum values of d and x, we can determine the maximum value of i for different m as follows:

$$i \le m - \lfloor m/P \rfloor \tag{B.3.10}$$

and

$$iP \le m(P-1) + < m >_P < N \tag{B.3.11}$$

since $N > (P^2+1)$. Consequently, given that $\tilde{h}(x,y) = 0$ for $x, y > P\text{-}1$ as specified in Eq. (B.3.5), we have

$$\tilde{h}(x,< d+mx-nP >_N) = \begin{cases} h(x,< d+m-iP >_N) & \text{if} \quad n=i \\ 0 & \text{if} \quad n \neq i \end{cases}$$

for $n = 0,\ldots, m-\lfloor m/P \rfloor$. Thus, we have

$$\sum_{n=0}^{m-\lfloor m/P \rfloor} \tilde{h}(x,< d+mx-nP >_N) = h(x,d+mx-iP) = h(x,< d+mx >_P)$$

Lemma 3.2 case (ii) is thus proved. (Q.E.D.)

Lemma 3.2 case (iii) can be proved similarly. We need to prove that

$$\sum_{n=0}^{m-\lfloor m/P \rfloor} \tilde{h}(x,< d+(N-1-m)x+nP >_N) = h(x,< d+(P-1-m)x >_P)$$

for $d, x = 0,\ldots, P\text{-}1$; $m = 0,\ldots, (P\text{-}1)/2$. The maximum value of the term $d+(P\text{-}1\text{-}m)x$ can be determined to be:

$$d+(P-1-m)x \le P-1+(P-1)(P-1) \le (P^2-P) < N$$

Hence, since both d, x and m are all positive integers, the term $h(x,< d+(P-1-m)x >_P)$ can be rewritten as follows:

$$\begin{aligned} h(x,< d+(P-1-m)x >_P) &= h(x,< d+Px-x-mx-iP >_N) \\ &= h(x,< d+(N-1-m)x+(x-i)P >_N) \\ &= h(x,< d+(N-1-m)x+i'P >_N) \end{aligned} \qquad \text{(B.3.12)}$$

for an integer i. From the minimum value of d and the maximum value of x, we can determine the maximum value of i' for different m as follows:

$$i' \le m-\lfloor m/P \rfloor \qquad \text{(B.3.13)}$$

and

$$i'P \le m(P-1)+< m >_P < N \qquad \text{(B.3.14)}$$

since $N > (P^2+1)$. Consequently, given that $\tilde{h}(x,y) = 0$ for $x, y > P-1$ as specified in Eq. (B.3.5), we have

$$\tilde{h}(x,< d+(N-1-m)x+nP >_N) = \begin{cases} h(x,< d+(N-1-m)x+iP >_N) & \text{if} \quad n=i \\ 0 & \text{if} \quad n \neq i \end{cases}$$

for $n = 0,\ldots, m-\lfloor m/P \rfloor$. Thus, we have

$$\begin{aligned} \sum_{n=0}^{m-\lfloor m/P \rfloor} \tilde{h}(x,< d+(N-1-m)x+nP >_N) &= h(x,d+(N-1-m)x+iP) \\ &= h(x,< d+(N-1-m)x >_P) \end{aligned}$$

for $d, x = 0,\ldots, P-1$; $m = 0,\ldots, (P-1)/2$. Lemma 3.2 case (iii) is thus proved.

(Q.E.D.)

Appendix C

Proof of Lemma 3.3

Given that a function $\tilde{h}$ with size $N \times N$ is formed by padding appropriate number of zeros to another function h with size $P \times P$, where $N = 2^n$; $n \in \mathbf{Z}$ and P is a prime integer. Both $\tilde{h}$ and h are assumed to be real functions. Assume also that the DPRT of $\tilde{h}$ is $\{\tilde{h}_m^c(d), \tilde{h}_s^b(d)\}$, where $m, d = 0,\ldots, N\text{-}1$; $s = 0,\ldots, N/2\text{-}1$, and the DPRT of h is $\{h_m^c(d), h_0^b(d)\}$, where $m, d = 0,\ldots, P\text{-}1$. Let the FFT of $\tilde{h}$ be $\tilde{H}$ and the FFT of h be H. Furthermore, let the FFT of $\{\tilde{h}_m^c, \tilde{h}_s^b\}$ be $\{\tilde{H}_m^c, \tilde{H}_s^b\}$. It is given that h is symmetric such that $h(x,y) = h(<P-x-1>_P, <P-y-1>_P)$.

(1) If h is symmetric and real, $\tilde{h}$ is also symmetric and real. Hence $\tilde{H}$ will be symmetric in a sense that

$$\tilde{H}(u,v) = \tilde{H}^*(<N-u>_N, <N-v>_N)$$

where $u, v = 0,1,\ldots, N\text{-}1$.

where $\tilde{H}^*$ is the complex conjugate of $\tilde{H}$. Consider the discrete Fourier slice theorem [75] as:

$$\tilde{H}_s^b(u) = \tilde{H}(u, <-2su>_N) = \sum_{d=0}^{N-1} \tilde{h}_s^b(d) \exp(-j2\pi ud/N) \tag{C.3.1}$$

$$\tilde{H}_m^c(v) = \tilde{H}(<-mv>_N, v) = \sum_{d=0}^{N-1} \tilde{h}_m^c(d) \exp(-j2\pi vd/N) \tag{C.3.2}$$

where $m = 0,1,\ldots, N\text{-}1$; $s = 0,1,\ldots, N/2\text{-}1$. Since

$\tilde{H}(u,v) = \tilde{H}^*(<N-u>_N, <N-v>_N)$, it implies that

$$\begin{aligned} \tilde{H}(u, <-2su>_N) &= \tilde{H}^*(<N-u>_N, <N-(-2su)>_N) \\ &= \tilde{H}^*(<N-u>_N, <-2s(N-u)>_N) \end{aligned} \tag{C.3.3}$$

$$\begin{aligned}\widetilde{H}(<-mv>_N, v) &= \widetilde{H}(<N-(-mv)>_N, <N-v>_N) \\ &= \widetilde{H}(<-m(N-v)>_N, <N-v>_N)\end{aligned} \tag{C.4}$$

Eqs. (C.3.3) and (C.3.4) show that the FFTs of $\{\widetilde{h}_m^c(d), \widetilde{h}_s^b(d)\}$ are also symmetric. Hence $\{\widetilde{h}_m^c(d), \widetilde{h}_s^b(d)\}$ must also be symmetric. However, since the imaginary part of $\{\widetilde{H}_m^c(d), \widetilde{H}_s^b(d)\}$ may not be equal to zero, $\{\widetilde{H}_m^c(d), \widetilde{H}_s^b(d)\}$ is only symmetric up to a rotation.

(Q.E.D.)

(2) We know that

$$\widetilde{h}_0^c(d) = \sum_{x=0}^{N-1} \widetilde{h}(x,d) = \begin{cases} \sum_{x=0}^{P-1} h(x,d) & \text{for } d = 0,...,P-1 \\ 0 & \text{for } N > d > P-1 \end{cases} \tag{C.3.5}$$

$$\widetilde{h}_0^b(d) = \sum_{y=0}^{N-1} \widetilde{h}(d,y) = \begin{cases} \sum_{y=0}^{P-1} h(d,y) & \text{for } d = 0,...,P-1 \\ 0 & \text{for } N > d > P-1 \end{cases} \tag{C.3.6}$$

If h is symmetric, $h(x,y) = h(y,x)$. Hence

$$\sum_{x=0}^{P-1} h(x,d) = \sum_{y=0}^{P-1} h(d,y)$$

This implies that $\widetilde{h}_0^c(d) = \widetilde{h}_0^b(d)$. (Q.E.D.)

(3) If h is real and symmetric such that $h(x,y) = h(<P-x-1>_P, <P-y-1>_P)$, $\widetilde{H}$ will also be symmetric in a sense that

$$abs(\widetilde{H}(u,v)) = abs(\widetilde{H}(<N-u>_N, v)) \tag{C.3.7}$$

$$abs(\widetilde{H}(u,v)) = abs(\widetilde{H}(u, <N-v>_N)) \tag{C.3.8}$$

If the discrete Fourier slice theorem is considered:

$$\widetilde{H}_s^b(u) = \widetilde{H}(u, <-2su>_N) = \sum_{d=0}^{N-1} \widetilde{h}_s^b(d) \exp(-j2\pi ud/N) \tag{C.3.9}$$

$$\tilde{H}_m^c(v) = \tilde{H}(<-mv>_N, v) = \sum_{d=0}^{N-1} \tilde{h}_m^c(d) \exp(-j2\pi vd/N) \tag{C.3.10}$$

where m = 0,1,..., N-1; s = 0,1,..., N/2-1. Eqs. (C.3.7) and (C.3.8) imply that

$$\begin{aligned} abs(\tilde{H}(u, <-2su>_N)) &= abs(\tilde{H}(u, <N-(-2su)>_N)) \\ &= abs(\tilde{H}(u, <-2(N/2-s)u>_N)) \end{aligned} \tag{C.3.11}$$

$$\begin{aligned} abs(\tilde{H}(<-mv>_N, v)) &= abs(\tilde{H}(<N-(-mv)>_N, v)) \\ &= abs(\tilde{H}(<-(N-m)v>_N, v)) \end{aligned} \tag{C.3.12}$$

Eqs. (C.3.11) and (C.3.12) show that

$$abs\left(FFT\left\{\tilde{h}_m^c(d), \tilde{h}_s^b(d)\right\}\right) = abs\left(FFT\left\{\tilde{h}_{N-m}^c(d), \tilde{h}_{N/2-s}^b(d)\right\}\right)$$

where m = 1,..., N-1; s = 1,..., N/2-1.

This implies the coefficients of projections (m, s) and (N-m, N/2-s) are identical up to a rotation.

(Q.E.D.)

Chapter 4 Improved MPEG-4 Still Texture Codec for Image Denoising and Compression

4.1 Introduction

In the previous chapter, we have proposed an efficient algorithm using DPRT for blind blurred image restoration. However, in some cases, noise introduced by the camera itself and the capturing environment (background noise) can also be the main sources of degradation for the acquired images and video frames. When coding that kind of noisy images, the performance of most image coding schemes can be greatly degraded. The recently proposed MPEG-4 still texture image codec is no exception.

The MPEG-4 standard adopted the zerotree concept, which is similar to the embedded zerotree wavelet (EZW) coding algorithm [9, 10], to serialize the wavelet coefficients of an image into bits in order of importance for still texture image coding. It is fully embedded and can stop at any point to meet the bit rate constraint. It provides very high coding efficiency over wide range of bit rates and supports coding of arbitrarily shaped still texture objects [125]. However, for many practical image processing systems, particularly, the remote sensing, surveillance, or monitoring systems, image acquisitions are done in a very noisy outdoor environment. Due to many uncontrollable constraints, such as, aging of the camera, it is difficult to ensure the acquired images to be noise-free. In fact, for most imaging sensors, the associated sensor noise can be modeled either as Gaussian-

distributed or Poisson-distributed random process. Poisson-distributed noise only occurs when the image light level is extremely low and the photodetector possesses a large internal electron amplification. Most practical detector circuits are usually well modeled by assuming additive Gaussian noise dominated [126]. When coding the noisy images, noise invades the MPEG-4 still texture image encoder by introducing more large amplitude wavelet coefficients in detailed levels. Since the MPEG-4 still texture image encoder uses the zerotree scanning algorithm for the bitstream generation, the increase in large amplitude wavelet coefficients in detailed levels affects the algorithm in predicting the originally insignificant wavelet coefficients. It results in using extra bits for the coding of the noisy coefficients hence poor compression performance is expected. Actually, the problem of noisy source coding has been studied for many years. Some researchers studied the effect of noise to the quantization process and modeled the quantization error in order to predict the noise pattern after quantization [127, 128, 129] for further processing. Some researchers concentrated on giving an optimal codec design by considering the mean-square-error (MSE) of the reconstructed signal with respect to the original noise-free signal [130, 131, 132]. However, both approaches do not pay much attention to the actual visual quality of the decoded noisy images. It is well-known that, for general image coding, decoded image with the highest PSNR, comparing with the original noise-free image, do not necessarily give the best visual quality. We shall show that it is particularly the case for noisy image coding.

In this chapter, we show that, when using the MPEG-4 still texture image codec to compress a noisy image, the decoded image with the highest PSNR does not give the best visual quality but is degraded by the so-called "cross" shape artifact. We begin by analyzing the effect of acquisition noise on the performance of the MPEG-4 still texture image codec. We try to measure the distortion, both qualitatively and quantitatively, of the decoded image as compared with the original noise-free image. We study the criterion that affects the visual

quality of the decoded image and then propose an adaptive thresholding technique in order to solve the problem. Finally, an analytical study is performed for the selection and the validation of the threshold value used in the adaptive thresholding technique.

In the next section, the effect of acquisition noise on the performance of the MPEG-4 still texture image codec is presented. It is followed by a review of the wavelet shrinkage and hard thresholding techniques for image denoising. The proposed adaptive thresholding technique as well as the process to select the critical threshold are discussed. Simulation results of the proposed approach are then discussed and compared with the wavelet shrinkage/hard thresholding approaches.

4.2 Effect of Noise to the MPEG-4 Still Texture Image Codec

The basic structure of the MPEG-4 still texture image coding algorithm has been shown in Fig. 2.4. Its operations can be further illustrated in Fig. 4.1. As illustrated in the figure, the input image is first decomposed into coefficients of different subbands by the DWT. The coefficients are then quantized using an uniform midrise quantizer followed by predictive coding or zerotree scanning depending on which subband they belong to. Finally, arithmetic coding is applied and the resulted bitstream is then sent to the receiver.

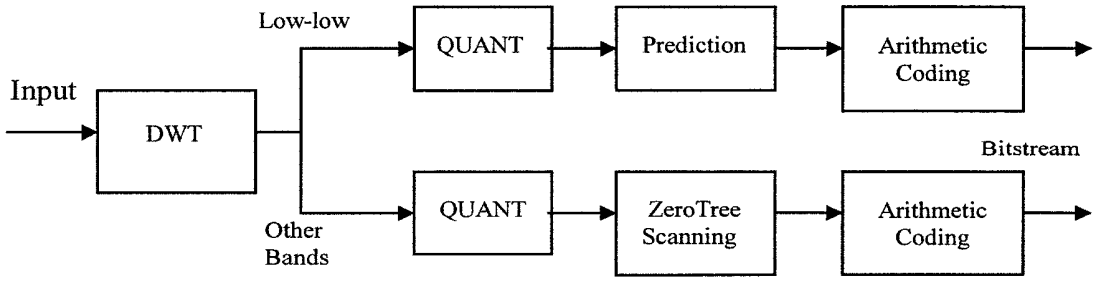


Figure 4.1: The MPEG-4 still texture image encoder.

We use an example to illustrate how noise introduced from the acquisition process affects the MPEG-4 still texture image codec. For simplicity, we assume the acquisition noise is additive and is of Gaussian distribution. In our example, a 256×256 image, Lenna, is first corrupted by additive zero mean Gaussian noise with noise variance $\sigma^2 = 10$ and 15, respectively. The noisy image is then encoded by the MPEG-4 still texture image codec. The PSNRs of the decoded noisy images with respect to the original noise-free image are studied. Fig. 4.2 shows the PSNR versus bit rate of the decoded noise-free Lenna and the decoded noisy Lenna with noise variance $\sigma^2 = 10$ and 15. The bit rates of the encoded image are achieved by adjusting the quantization value of the wavelet coefficients in higher subbands.

It is seen in Fig. 4.2 that the curves for the case of noisy image coding behave as that of the noise-free image coding in the low bit rate region and start to drop when bit rate increases. Both curves for the case of noisy image coding keep almost constant in the high bit rate region with the PSNR being close to the PSNR of the original noisy image.

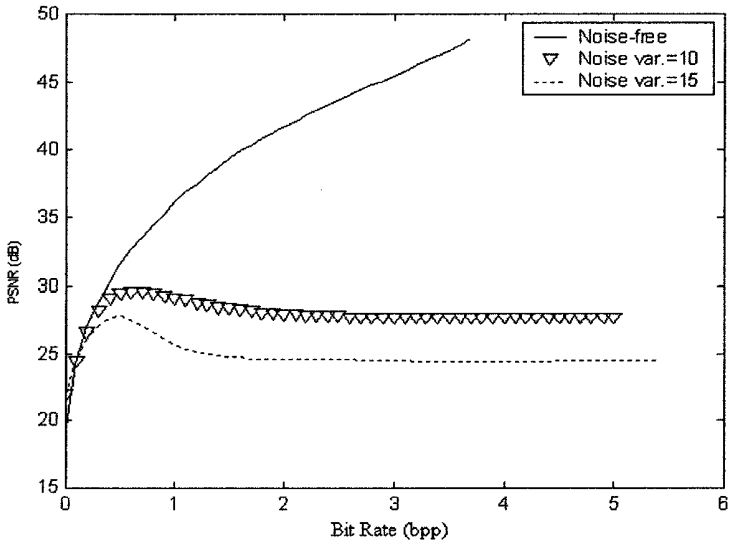


Figure 4.2: PSNR versus bit rate using MPEG-4 still texture image codec.

This is a typical result when coding general noisy images using the MPEG-4 still texture image codec. It is interesting to note that, first, for both curves under the case of noisy image coding, the highest PSNR appears in the low bit rate region. Reducing the compression rate cannot further improve PSNR but eventually makes it decrease to the PSNR of the original noisy image (to ease our later discussion, let us dub this point where the highest PSNR is obtained as the optimal operating point). Second, the shape of both curves for the case of noisy image coding is similar and the optimal operating point drifts to the lower bit rate region as noise variance increases. This interesting performance of the MPEG-4 still texture image codec can be explained as follows. Due to the linear property of the wavelet transform, the additive noise component in the image is still additive in the transform domain. As mentioned before, the noise components introduce more large amplitude wavelet coefficients in the detailed levels. As a result, the zerotree scanning cannot give accurate prediction of the originally insignificant wavelet coefficients. Additional resources have to be used to encode the extra detailed level coefficients. The compression rate is then greatly reduced. This explains why the performance of the decoded noisy images, in general, must be poorer than the decoded noise-free images. However, in the very low bit rate region, both curves for the cases of noise-free and noisy image coding give a similar shape as shown in Fig. 4.2. It is because, in this region, large quantization values are employed. Almost all coefficients in the higher subbands, as well as noise components, are quantized to zero, thus noise effect is negligible. This explains why it is possible that the PSNR of the decoded noisy image can be higher than that of the original noisy image.

The above result shows an interesting point that we can achieve the goal of denoising by merely performing quantization. Actually, when a large enough quantization value is applied, it is straightforward that most of the coefficients will be quantized to zero. Hence, a nearly noise-free quantized result can be obtained. In general, by considering the

absolute amplitude of the signal, we can achieve denoising to a certain extent by merely quantization if the following conditions are satisfied:

$$\begin{cases} (x \bmod Q) + n < Q \\ n < Q \end{cases} \tag{4.1}$$

where x is the absolute amplitude of the original signal; 'mod' denotes the modulus operator; n represents the noise component and Q is the quantization value. Mathematically, once Eq. (4.1) is satisfied, the quantization process of a noisy signal x can be modeled as:

$$\left\lfloor \frac{x+n}{Q} \right\rfloor = \left\lfloor \frac{x}{Q} \right\rfloor + \left\lfloor \frac{n}{Q} \right\rfloor.$$

Since $\left\lfloor \frac{n}{Q} \right\rfloor = 0$ as implied in Eq. (4.1), hence

$$\left\lfloor \frac{x+n}{Q} \right\rfloor = \left\lfloor \frac{x}{Q} \right\rfloor \tag{4.2}$$

where $\lfloor \cdot \rfloor$ denotes the floor operator which rounds the elements to the nearest integers towards minus infinity. Eq. (4.2) shows that the noise component does not show any effect to the original signal x after quantization. However, only part of the signals may satisfy Eq. (4.1) during the quantization process and hence denoising can only be achieved to a certain extent.

When the quantization value decreases, the ability of the quantizer in removing noise will decrease. However, the quantization error will also decrease. There will be a point where the error generated due to quantization and the acquisition noise is the minimum, of which we refer as the optimal operating point of the coder. At this point, the decoded image with the best quality, measured in terms of PSNR, can be obtained theoretically.

Yet, it is well-known in general image coding that decoded image with the highest PSNR does not necessarily imply the best visual quality. It is particularly the case for noisy

image coding. Fig. 4.3 shows the reconstructed images encoded at different regions of the resulting curve at $\sigma^2 = 10$ as shown Fig. 4.2. The decoded image shown in Fig. 4.3(b), encoded at bit rate = 0.599 bpp has the highest PSNR. It is in fact the optimal operating point of the coder at $\sigma^2 = 10$. However, within the smooth region of that image, there exists sharp "cross" shape artifact. This artifact appears in the coding of all general noisy images using the MPEG-4 still texture image codec and it greatly affects the visual quality of the decoded images.

There are basically three ways to solve the problem. First, we can introduce an extra denoising step before the coding stage, but it implies extra computational effort. Second, we can embed the denoising step into the codec. Since the MPEG-4 still texture image coder uses DWT, the wavelet denoising methods, such as wavelet shrinkage or hard thresholding, can be applied to the transformed coefficients. However, in this chapter, we propose the third method, namely adaptive thresholding, which outperforms the traditional wavelet denoising approaches. The improvement is mainly due to the consideration of the source of such "cross" shape artifact when designing the algorithm.

In the next section, we review the wavelet shrinkage and the hard thresholding approaches for noise reduction. After that, we investigate the reason why this kind of "cross" shape noise appears in the decoded images. We then propose an adaptive thresholding technique for the quantization process of the MPEG-4 still texture image coder to remove the "cross" shape artifact in the decoded images. This approach allows us to encode images at the optimal operating point without the degradation in visual quality. Hence, both PSNR and the visual quality of the decoded noisy images are optimized.

(a)

(b)

(c)

Figure 4.3: Decoded image with noise variance σ^2 = 10 at (a) bit rate = 0.263 bpp and PSNR = 27.62 dB; (b) bit rate = 0.599 bpp and PSNR = 29.54 dB (optimal case); (c) bit rate = 2.730 bpp and PSNR = 27.75 dB.

4.3 Review on Wavelet Shrinkage and Hard Thresholding

As mentioned in the previous section, noise components greatly reduce the compression efficiency of the MPEG-4 still texture image coding scheme as well as the quality of its reconstructed images. Therefore, pre-processing step for reducing noise effect before the actual encoding process is necessary. There are many approaches proposed for noise reduction as discussed in Section 1.3.2. In particular, the wavelet-based shrinkage and the hard thresholding approaches have aroused much attention due to their simplicity and effectiveness.

In this section, we shall review the wavelet-based noise reduction methods including wavelet shrinking and hard thresholding. For simplicity, we use 1-D signal for elaboration. Suppose we wish to recover an unknown 1-D signal f from its noisy observation y such that:

$$y_i = f_i + n_i \text{ for } i = 0, 1, \ldots, N\text{-}1 \tag{4.3}$$

where n is an independent and identically distributed white Gaussian noise with zero mean and noise variance σ^2 and N is the length of the signal, respectively. Let $\hat{f}$ be the estimate of f. The goals of shrinkage and hard thresholding are to optimize the estimate by minimizing the following MSE:

$$\sum_{i=0}^{N-1} E\left[f_i - \hat{f}_i\right]^2. \tag{4.4}$$

The simple wavelet domain thresholding method has three steps:

1. Compute the orthonormal DWT of the noisy data y to obtain its corresponding scaling and wavelet coefficients.
2. The scaling coefficients are left unchanged in order that gross structure is not lost. To each individual wavelet coefficient, apply a threshold rule, either soft (a continuous function of the data which shrinks each observation, also called shrinkage) or hard (which retains only large observations) using the following formulation:

$$n_s(w,t) = \begin{cases} w-t, & \text{for} \quad w \geq t \\ 0, & \text{for} \quad |w| < t \\ w+t, & \text{for} \quad w \leq -t \end{cases} \tag{4.5}$$

$$n_h(w,t) = \begin{cases} w, & \text{for} \quad |\mathrm{w}| \geq t \\ 0, & \text{for} \quad |\mathrm{w}| < t \end{cases} \tag{4.6}$$

where w is the wavelet coefficient and t is a threshold equals to $\sigma\sqrt{2\log(N)}$, and $n_s(w,t)$ and $n_h(w,t)$ represent the shrinkage and hard thresholding operators, respectively.

3. Perform the inverse orthonormal DWT to obtain the estimate $\hat{f}$ using the scaling coefficients and the thresholded wavelet coefficients.

The threshold t here is called universal threshold. It is designed for suppressing noise-induced spikes which spoil the smoothness of reconstruction. The choice of $t = \sigma\sqrt{2\log(N)}$ is based on the proof that the maximum of N standard Gaussian variates having variance σ^2 is smaller than $\sigma\sqrt{2\log(N)}$ with probability increasing to 1 as N increases. One important qualitative feature of this method is that the relatively sharp features in f, for example, the edges, are maintained while the noise is suppressed [54].

It should be noted that the shrinkage is the MSE optimal nonlinear function to apply in the orthonormal wavelet domain if one requires the reconstructed signal to be at least as smooth as the original noise-free signal. In fact, the denoised signal using the hard thresholding preserves features, for example, peak heights, better, but it may exhibit greater spurious oscillations in the vicinity of discontinuities than that using the shrinkage.

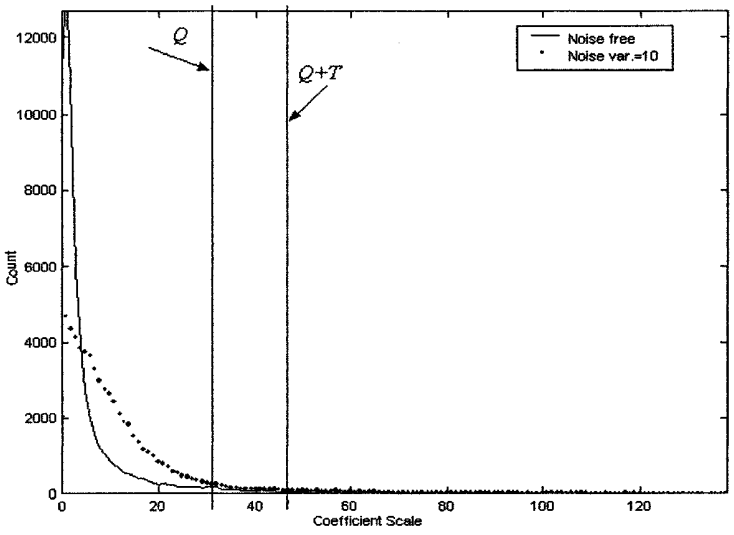
4.4 Proposed Adaptive Thresholding Technique

4.4.1 Motivation

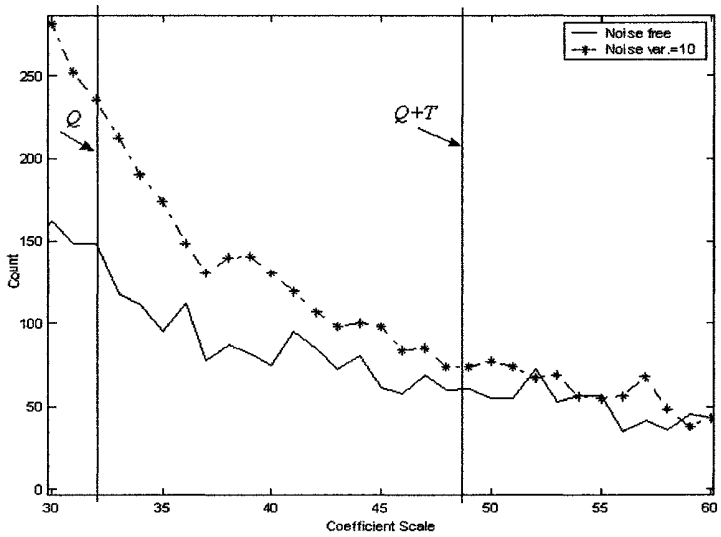
As shown in the Section 4.3, the visual quality of the reconstructed noisy image is often degraded by the so-called "cross" shape artifact. This "cross" shape artifact is due to the residual noise impulse generated during the quantization process of the encoder [133]. For low bit rate coding, it is mentioned in Section 4.2 that some of the noise components, which fulfill the condition as stated in Eq. (4.1), is removed in the quantization process. However, it is possible that some wavelet coefficients will falsely jump from one quantization level to another due to the additive noise. These suddenly jumped wavelet coefficients, which violate the condition set in Eq. (4.1), will introduce impulse-like noises to the final bitstream after quantization. After the inverse 2-D DWT performed in the decoder, the impulse-like noise becomes the "cross" shape artifact in the reconstructed images.

To deal with this problem, we propose an adaptive thresholding technique to remove the "cross" shape artifact by reducing the falsely jumped wavelet coefficients. It should be noted that the optimal operating point often appears in the low bit rate region where the quantization step value is very large. For this region, most wavelet coefficients locate in the first and second quantum levels. It implies that most falsely jumped wavelet coefficients locate in these two quantum levels. To effectively reduce the falsely jumped wavelet coefficients, we can increase the step value of the first quantum by a threshold T. By doing so, the wavelet coefficients, which falsely jumped from the first quantum to the second quantum, will be forced back to the first quantum. Without losing the generality, we use the image Lenna again to illustrate the proposed approach. Fig. 4.4(a) shows the histograms of the wavelet coefficients of the original and noisy Lenna after DWT. From Fig. 4.4(a), the distribution of the noisy wavelet coefficients spreads over a larger amplitude range as compared with the noise-free case. The spreading of noisy wavelet coefficients is mainly

due to the noise effect. If the quantization step value for the first quantum level is set at Q = 32 as shown in Fig. 4.4(a), many of the quantized data in the second quantum level, in fact, are due to the falsely jumped wavelet coefficients. This can be observed by comparing the histograms of the noise-free and the noisy wavelet coefficients in Fig. 4.4(b). Now if we increase the step value to $Q+T$, where T is a threshold, as shown in Fig. 4.4(b), most falsely jumped wavelet coefficients will then be forced back to the first quantum level. Nevertheless, the choice of this threshold value should be very careful. Although we can remove the falsely jumped wavelet coefficients on one side, some original wavelet coefficients are also removed on the other side. One reasonable criterion for choosing this threshold is to ensure the probability of noise being removed due to this threshold is greater than that of the original signal being eliminated. To do this, we have to analyze the distribution model of the wavelet coefficients.



(a)



(b)

Figure 4.4: (a) Histograms of the wavelet coefficients of the original and noisy Lenna at noise variance $\sigma^2 = 10$ after the DWT (absolute value) (b) Zoom-in of the curves at $Q = 32$.

4.4.2 Threshold selection

Let us extend the 1-D case in Eq. (4.3) to 2-D one for further elaboration. Suppose an image, f, which is obtained from a remote sensing system and is corrupted with zero mean white Gaussian noise, n. The received image Y can be described by the following equation:

$$Y(i,j) = f(i,j) + n(i,j) \quad \text{for } i,j = 0, 1,..., N\text{-}1 \tag{4.7}$$

where the sample size of f is N^2. Due to the fact that the wavelet transform maps white noise in the signal domain to white noise in the transform domain, the wavelet coefficients of a noisy signal are themselves just the noisy version of the noise-free wavelet coefficients [134]. Therefore, Eq. (4.7) can be rewritten by the following equation:

$$W_Y(i,j) = W_f(i,j) + n^*(i,j). \tag{4.8}$$

where $n^*(i,j)$ is the white Gaussian noise after DWT; $W_Y(i, j)$ and $W_f(i, j)$ are the wavelet coefficients of the noisy image $Y(i, j)$ and noise-free image $f(i,j)$, respectively. From Eq. (4.8), we obtain the noisy wavelet coefficient model by combining the original noise-free wavelet coefficient model and a Gaussian model. We suggest using the following two-parameter model to describe the distribution of the original noise-free wavelet coefficients. It has been proved to best fit the distribution of the wavelet coefficients of general images [134, 135]:

$$f(x) = Ke^{-(|x|/\alpha)^\beta}.$$

The parameters α and β represent the variance and the decreasing rate of the distribution of the detailed level wavelet coefficients, respectively; K represents the proportional constant and $f(.)$ is the distribution function of the wavelet coefficients in detailed levels. By using the above equation, the probability density function (pdf) of the wavelet coefficients can be approximately modeled as:

$$f_D(x) = \frac{f(x)}{N_x} = \frac{K}{N_x} e^{-(|x|/\alpha)^\beta} \tag{4.9}$$

where N_x is the total number of wavelet coefficients. In the MPEG-4 standard, all higher subbands wavelet coefficients are quantized by an uniform midrise quantizer and only the absolute amplitude of the coefficient is considered while the sign of the coefficient is marked before [125]. The probability of noisy coefficients that falsely jumped from the first quantum $[0,Q)$ to the second quantum $[Q, 2Q)$ can be approximately modeled as:

$$
\begin{aligned}
&P\Big(Q \le |x+n^*| < 2Q \;\Big|\; |x| < Q \Big) \\
&= 2\lim_{a\to Q}\left\{ \int_0^a f_D(x) \int_{a-x}^{2a-x} g_\sigma(n)\,dn\,dx + \int_0^a f_D(x) \int_{-a-x}^{-2a-x} g_\sigma(n)\,dn\,dx \right\} \\
&\approx 2\lim_{a\to Q}\left\{ \int_0^a f_D(x) \int_{a-x}^{2a-x} g_\sigma(n)\,dn\,dx \right\}
\end{aligned}
$$

where g_σ is the pdf of the zero mean white Gaussian noise with variance σ^2. The first term of the above equation corresponds to the case that x and n are of the same sign while the second term indicates the case when x and n are of opposite sign. As the probability for the occurance of the second term is low, we can then eliminate the second term. These falsely jumped wavelet coefficients are the main cause of the "cross" shape artifact. If we add a threshold T to the quantization step Q, the probability of the original signal having the amplitude smaller than Q, but falsely jumped to the range $[Q, Q+T)$ is:

$$
\begin{aligned}
&P\Big(Q \le |x+n^*| < Q+T \;\Big|\; |x| < Q \Big) \\
&\approx 2\lim_{a\to Q}\left\{ \int_0^a f_D(x) \int_{a-x}^{a+T-x} g_\sigma(n)\,dn\,dx \right\} \\
&\ge \lim_{a\to Q} \int_{-a}^{a} f_D(x)\,dx \int_a^{a+T} g_\sigma(n)\,dn
\end{aligned}
\tag{4.10}
$$

The inequality in Eq. (4.10) is valid since for zero mean Gaussian distribution function, g_σ,

$$
\int_A^B g_\sigma(n)\,dn \le \int_{A-\varepsilon}^{B-\varepsilon} g_\sigma(n)\,dn
$$

where A, B, and ε are positive numbers and $\varepsilon < A, B$. By setting all coefficients having the amplitudes smaller than $Q+T$ to zero (that is, back to quantum level 1), a great amount of

noisy coefficients, which is directly proportional to $P\left(Q \le |x+n^*| < Q+T \mid |x| < Q\right)$, will be recovered. However, as mentioned before, if all the coefficients having the value smaller than $Q+T$ are set to zero, some wavelet coefficients which belong to the range $[Q, Q+T)$ originally will be removed as well. We should carefully select the threshold T such that the probability to remove noise is higher than that of the original signal in the range $[Q, Q+T)$. To evaluate the distribution of the coefficients in the range $[Q, Q+T)$, let us consider their probability as given by the following formulation:

$$\begin{aligned} & P\left(Q \le |x+n^*| < Q+T \mid Q \le |x| < Q+T \right) \\ & \approx 2\lim_{a \to Q}\left\{ \int_a^{a+T} f_D(x) \int_{a-x}^{a+T-x} g_\sigma(n)\,dn\,dx \right\} \\ & \le 2\lim_{a \to Q}\left\{ \int_a^{a+T} f_D(x)\,dx \int_{-T/2}^{T/2} g_\sigma(n)\,dn \right\} \\ & \le \left[1 - \lim_{a \to Q} \int_{-a}^{a} f_D(x)\,dx \right] \int_{-T/2}^{T/2} g_\sigma(n)\,dn \end{aligned} \tag{4.11}$$

The inequality in Eq. (4.11) is valid since, it can be shown that, for zero mean Gaussian distribution, g_σ,

$$\int_{-T/2}^{T/2} g_\sigma(n)\,dn \ge \int_A^{A+T} g_\sigma(n)\,dn$$

where A and T are any real numbers. If all the data in $[Q, Q+T)$ are removed, the probability of the original noise-free wavelet coefficients within this range after quantization is directly proportional to $P\left(Q \le |x+n^*| < Q+T \mid Q \le |x| < Q+T \right)$. By comparing Eqs. (4.10) and (4.11), we know that if we have the following relationship:

$$\lim_{a \to Q} \int_{-a}^{a} f_D(x)\,dx \int_a^{a+T} g_\sigma(n)\,dn > \left[1 - \lim_{a \to Q} \int_{-a}^{a} f_D(x)\,dx \right] \cdot \int_{-T/2}^{T/2} g_\sigma(n)\,dn \tag{4.12}$$

then,

$$P\left(Q \le |x+n^*| < Q+T \| x| < Q\right) > P\left(Q \le |x+n^*| < Q+T | Q \le |x| < Q+T\right).$$

That is, the probability to remove the falsely jumped coefficients is higher than that of the original coefficients in the range [Q, Q+T). Eq. (4.12) can be rewritten as:

$$\lim_{a\to Q}\int_{-a}^{a} f_D(x)dx\left[\int_{a}^{a+T} g_\sigma(n)dn+\int_{-T/2}^{T/2} g_\sigma(n)dn\right] > \int_{-T/2}^{T/2} g_\sigma(n)dn$$

$$\lim_{a\to Q}\int_{-a}^{a} f_D(x)dx > \frac{\int_{-T/2}^{T/2} g_\sigma(n)dn}{\left[\lim_{a\to Q}\int_{a}^{a+T} g_\sigma(n)dn+\int_{-T/2}^{T/2} g_\sigma(n)dn\right]}$$

since

$$\int_{-T/2}^{T/2} g_\sigma(n)dn \le \frac{\int_{-T/2}^{T/2} g_\sigma(n)dn}{\left[\lim_{a\to Q}\int_{a}^{a+T} g_\sigma(n)dn+\int_{-T/2}^{T/2} g_\sigma(n)dn\right]}.$$

Thus, if we have

$$\lim_{a\to Q}\int_{-a}^{a} f_D(x)dx > \int_{-T/2}^{T/2} g_\sigma(n)dn. \tag{4.13}$$

Eq. (4.12) will be valid. The left hand side of Eq. (4.13) can be considered as the cumulative probability of wavelet coefficients which have amplitudes smaller than the quantization step value Q. The right hand side of Eq. (4.13) represents the cumulative probability under the Gaussian distribution due to the threshold T. Our objective here is that, for a given quantization step value Q and a given noise variance σ^2, we try to determine the largest threshold value T that fulfills Eq. (4.13). This will ensure that the probability to remove the falsely jumped coefficients is higher than that of the original coefficients in the range $[Q, Q+T)$. To achieve our objective, we directly compare the cumulative probability under the wavelet coefficient distribution and the Gaussian distribution due to the quantization values applied and the thresholds used, respectively.

By using a typical wavelet coefficient distribution model, let us determine the cumulative probability of wavelet coefficients which have amplitudes smaller than the quantization step value Q for general images. Fig. 4.5 shows the wavelet coefficient

distributions for four levels of decomposition of five standard images, including Lenna, Peppers, House, Bird, and Bridge. It is seen that these distributions are so similar that it is possible to represent them using a concrete model obtained by simply averaging these distributions. Fig. 4.6 shows the resulting distribution obtained by averaging those obtained from 9 standard images, including Lenna, Peppers, House, Baboon, Bird, Bridge, Camera man, Circle, and Coifman. We also show in Fig. 4.6 the curve obtained by fitting the model suggested in Eq. (4.9) to the averaged distribution obtained above. It can be seen that these two curves are very close. This result validates the accuracy of the model suggested in Eq. (4.9) in representing the wavelet coefficient distribution for general images.

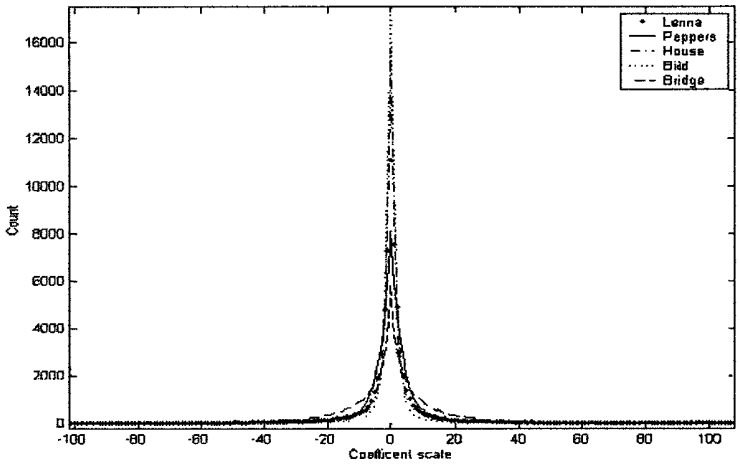


Figure 4.5: Distributions of detailed level wavelet coefficients for Lenna, Peppers, House, Bird, and Bridge (for four levels of decomposition).

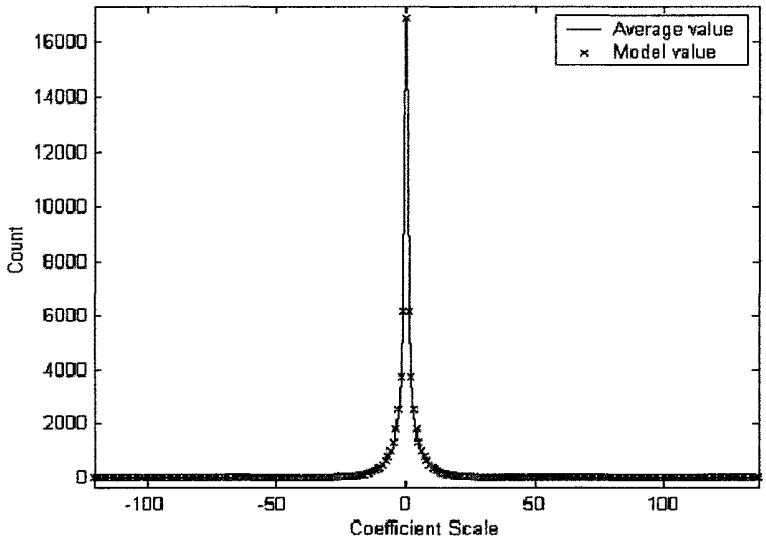


Figure 4.6: The average distribution of the wavelet coefficients for 9 standard images and the curve generated by the model given in Eq. (4.9) with $K = 1.6906\times10^4$, $\alpha = 0.980$ and $\beta = 0.575$. It shows that they are very similar.

By using the average distribution as shown in Fig. 4.6, the relationship between the quantization step value and the cumulative probability (due to that quantization step) under the wavelet coefficient distribution can be established as shown in Fig. 4.7. For instance, if the quantization step value $Q = 32$, the cumulative probability of wavelet coefficients which have amplitudes smaller than Q is about 91%.

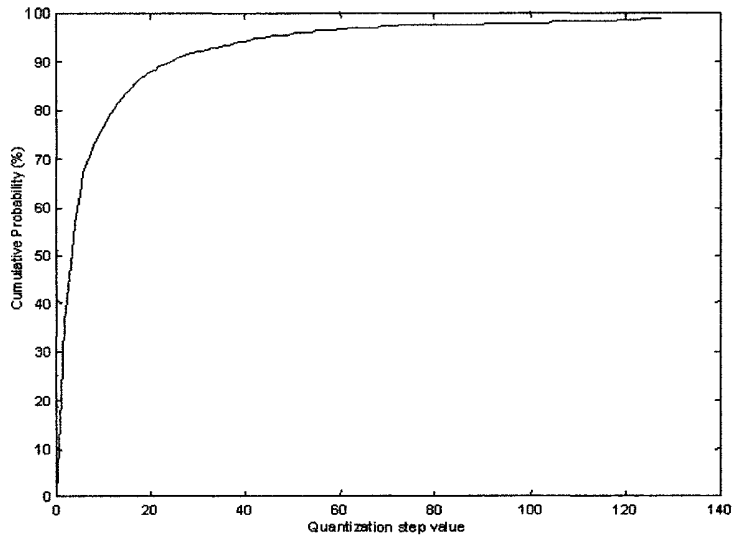


Figure 4.7: The relationship between the quantization step value and the cumulative probability of wavelet coefficients which have magnitudes smaller than Q. This graph is derived based on the average distribution suggested in Fig. 4.6.

On the other hand, when considering the cumulative probability under the Gaussian distribution due to the threshold T, it should first be noted that the Gaussian distribution is a function of the noise variance σ^2. For different noise variances, different threshold T should be chosen in order to fulfill Eq. (4.13). However, it is well-known that the cumulative probability under the zero mean Gaussian distribution for all variance σ^2 is constant if we evaluate it in the range $[-k^*\sigma, k^*\sigma]$, where k^* is a constant. It gives us a simple way to relate the threshold T with the noise variance of the image. More specifically, let us simply express T as a function of σ, i.e., $T = 2k^*\sigma$. Fig. 4.8 shows a plot of the percentage of the cumulative probability under the zero-mean Gaussian distribution in the range $[-k^*\sigma, k^*\sigma]$. The result in Fig. 4.8 can be compared with that in Fig. 4.7 to determine the appropriate T with known Q and σ^2. The proposed algorithm can be summarized into the following steps:

1. For a particular Q, we look for the corresponding cumulative probability from Fig. 4.7 for general images.

2. The estimated probability in step 1) is used to determine the required value k^* in Fig. 4.8 for zero-mean Gaussian distribution.
3. The threshold value T can be found by setting $T = 2k^*\sigma$ and the noise variance σ^2 can be found by a prior test on the image acquisition system using a testing image.

By adopting the above approach, we obtain the following set of threshold values T for typical quantization step values Q:

Quantization step Q	Threshold value T
2	0.936σ
4	1.544σ
8	2.16σ
16	2.80σ
32	3.52σ
64	4.10σ
128	5.10σ

Table 4.1: Threshold values correspond to different quantization step values.

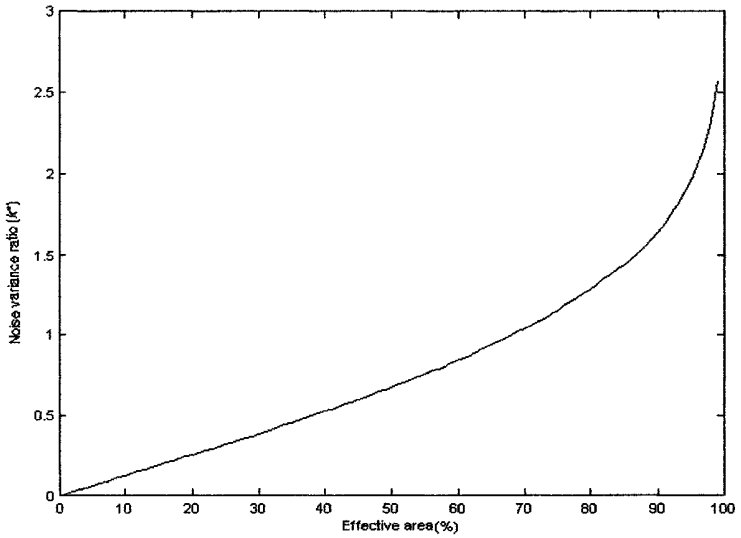


Figure 4.8: Noise variance ratio (k^*) against the cumulative probability under the zero-mean Gaussian distribution.

4.5 Simulation Results

It should be noted that the proposed approach is different from other generic denoising algorithms, such as, the wavelet shrinkage and low-pass filtering. For generic denoising algorithms, they take no consideration on the quantization problem of the encoder such that they generally cannot remove the "cross" shape artifact, which is the major problem of coding noisy images. We have compared our proposed approach with other generic wavelet-based techniques such as the wavelet shrinkage and hard thresholding techniques as discussed in Section 4.3 using the universal threshold. Besides, the well-known simplest low-pass filtering approach is also cited for our comparison. For low-pass filtering, all the wavelet coefficients located in the highest subbands are removed in order to filter out the noise. Single quantization mode is applied for all the simulations. Fig. 4.9 shows a comparison of the results obtained by using the above approaches in coding a noisy image, Lenna, with the MPEG-4 still texture image coder.

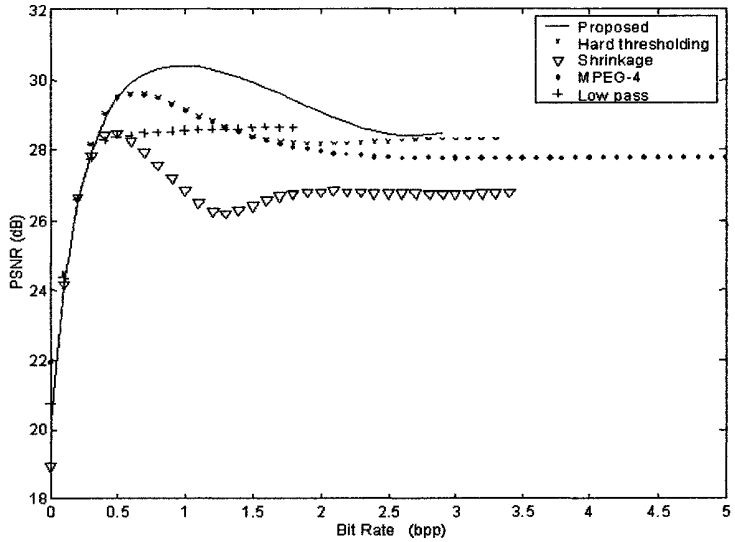


Figure 4.9: Comparison of PSNR against bit rate for noisy Lenna at noise variance $\sigma^2 = 10$, by using the proposed approach, hard thresholding, wavelet shrinkage, traditional MPEG-4 and the low-pass filtering approach.

From Fig. 4.9, it can be shown that the proposed approach gives the best performance among the other 3 mentioned approaches. It can be shown from Fig. 4.9 that the wavelet shrinkage approach performs even poorer than the MPEG-4 approach without denoising. It is because the wavelet shrinkage approach can only guarantee to provide the smoothest denoised signal but not the highest resulted PSNR. The resulted PSNRs of the decoded noisy images are improved to a great extent, particularly in the low bit rate region (which is our main concern). Figs. 4.10-4.13, show the decoded images by using the hard thresholding, the wavelet shrinkage, the low-pass filtering, and the proposed adaptive thresholding approach with similar compression ratio.

Figure 4.10: Decoded image by hard thresholding bit rate=0.51 bpp, PSNR = 29.47 at noise variance $\sigma^2 = 10$.

Figure 4.11: Decoded image by wavelet shrinkage at bit rate = 0.50 bpp, PSNR = 27.36, at noise variance $\sigma^2 = 10$.

Figure 4.12: Decoded image by low-pass filtering at bit rate = 0.50 bpp, PSNR = 28.35, at noise variance $\sigma^2 = 10$.

Figure 4.13: Decoded image by the proposed approach at bit rate = 0.49 bpp, PSNR = 29.46, at noise variance $\sigma^2 = 10$.

From Figs. 4.10 and 4.11, it can be easily seen that, the decoded images using the hard thresholding and the wavelet shrinkage are affected by the "cross" shape noise that can be found in the smooth regions of the images. They degrade the visual quality of the decoded images to a great extent. For the low-pass filtering approach, it can be seen in Fig. 4.12 that, the decoded image becomes blur due to the removal of the detailed level wavelet coefficients. It is seen in Fig. 4.13 that, when using the proposed approach in MPEG-4 still texture image coding, the "cross" shape noise is greatly reduced while keeping reasonable good compression ratio as well as PSNR. Fig. 4.14 shows the performances of the proposed adaptive thresholding approach together with other denoising approaches when noise variance σ^2 is increased to 15. From Fig. 4.14, it is shown that similar results can be obtained when noise variance increases. The unbiased results of the proposed approach ascertain its background theory and the stability.

The adaptive thresholding technique has been applied to the coding of other standard images, such as Peppers. Figs. 4.15 and 4.16 show the performances of the proposed approach, together with other denoising approaches, in coding noisy Peppers with the

MPEG-4 still texture image coder at noise variance σ^2 = 10 and 15. The results show that, while improving the visual quality of the decoded images to a great extent, a 1 to 2 dB improvement in PSNRs is achieved when comparing with the traditional MPEG-4 still texture image coder with similar compression ratio. Besides, the proposed approach can be extended to other quantization styles, e.g. bi-level or multi-level quantization, easily as only the wavelet coefficients at the first scalability level have to be considered. More importantly, no extra denoising procedure is needed. The proposed technique requires only a slight modification to the MPEG-4 still texture image coder with the decoder keeping unchanged.

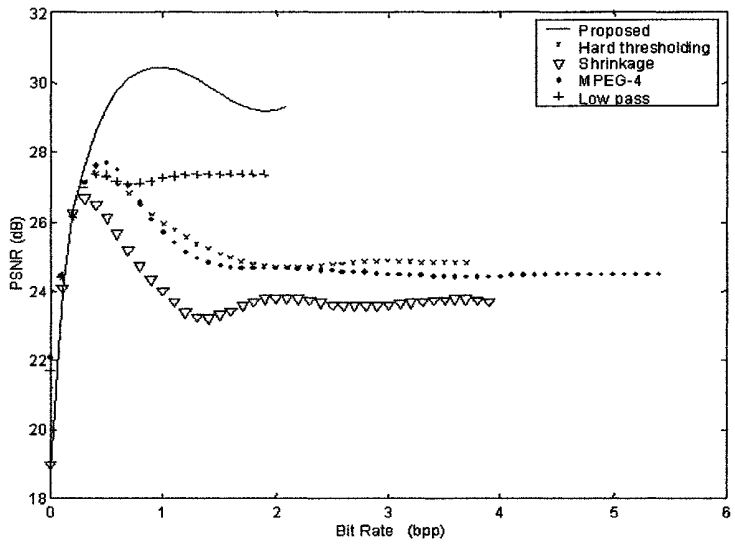


Figure 4.14: Comparison of PSNR against bit rate for noisy Lenna at noise variance σ^2 = 15, by using the proposed approach, hard thresholding, wavelet shrinkage, traditional MPEG-4, and the low-pass filtering approach.

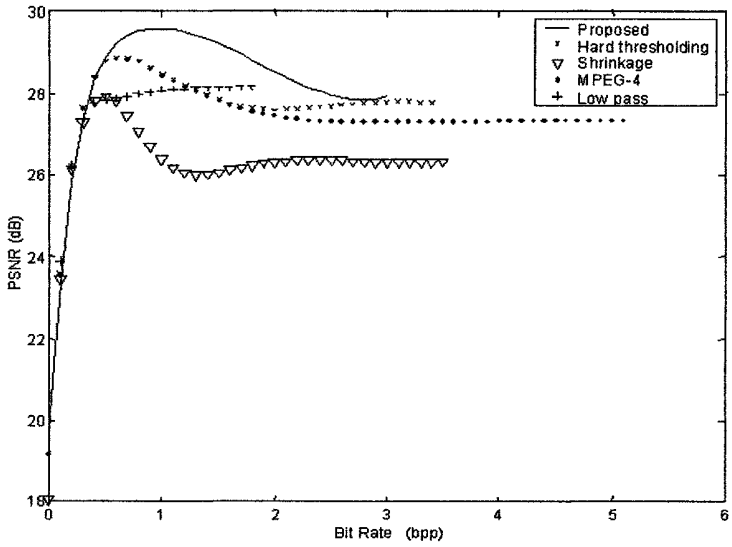


Figure 4.15: Comparison of PSNR against bit rate for noisy Peppers at noise variance $\sigma^2 = 10$, by using the proposed approach, hard thresholding, wavelet shrinkage, traditional MPEG-4, and the low-pass filtering approach.

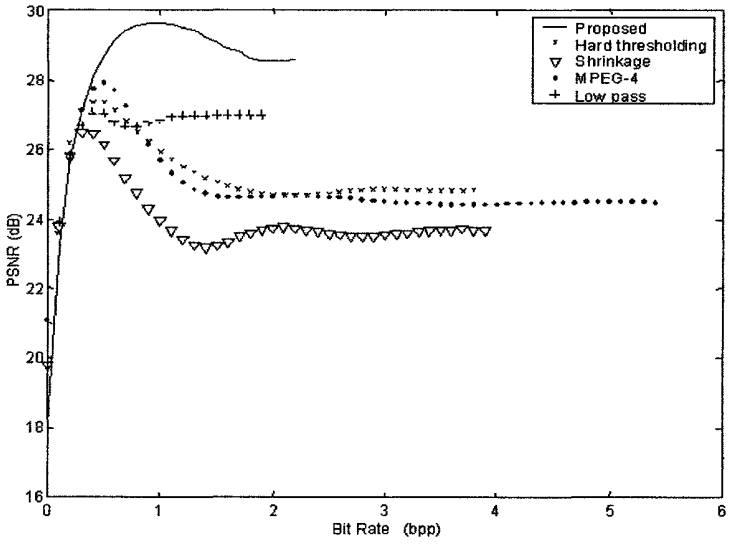


Figure 4.16: Comparison of PSNR against bit rate for noisy Peppers at noise variance $\sigma^2 = 15$, by using the proposed approach, hard thresholding, wavelet shrinkage, traditional MPEG-4, and the low-pass filtering approach.

4.6 Summary

In this chapter, we have studied and analyzed the effect of acquisition noise on the MPEG-4 still texture image coder. We show that the quantization process can partly remove the noise component in an image. However, the residual noise affects the coder in such a way that decreasing the compression rate may not necessarily imply improving the PSNR. This is the reason why an optimal operating point is noted when using the MPEG-4 still texture image coder to compress a noisy image. Nevertheless, the visual quality at the optimal point is greatly degraded by the "cross" shape artifact due to the quantization of the residual noise component. Then, a brief review on the wavelet shrinkage and the hard thresholding for noise reduction is given. Finally, we propose an adaptive thresholding technique for the MPEG-4 still texture image coder in order to remove that residual noise by considering the pdf of the wavelet coefficients. Results show that after applying the adaptive thresholding technique, the occurrence of "cross" shape noise in the reconstructed image is greatly reduced. The PSNRs of the decoded images are also improved by 1 to 2 dB as compared with the traditional MPEG-4 still texture image coder with similar compression ratio. It should be noted that the proposed technique requires only a slight modification to the coder while the decoder remains unchanged. The proposed technique is particularly suitable for the real time low bit rate image processing systems under noisy environment, such as, the remote monitoring system.

Chapter 5 Rate-Distortion (RD) Modeling for Image/Video Compression

5.1 Introduction

The rate-distortion (RD) theory was originally introduced by Shannon in [136] to explore the relationship between coding rate, R, and measured distortion, D, of a compressed source. It mainly concerns the task of representing a source with the fewest number of bits to achieve the best reproduction quality subject to a fidelity criterion. Here, the distortion D encounters the difference between the reconstruction and the original uncoded image/video source, which is often measured in terms of MSE.

To study the RD relationship of typical transform coding systems, different coding rates should be applied while other compression techniques, e.g. motion estimation and compensation, should be fixed. In such case, the factors affecting the RD performance are the quantization step, Q, and the group-of-picture (GOP) structure of the encoded sequence. Hence, the RD relationship can be formulated in terms of Q and decomposed into two characteristic functions, namely the rate-quantization, $R(Q)$, and the distortion-quantization, $D(Q)$, functions [137, 138]. By exploring the $R(Q)$ and $D(Q)$ functions, a specific Q can be determined to achieve target output bit rate or picture quality of an encoded source in order to ingratiate different system constraints [139, 140, 141, 142]. Furthermore, optimization can be performed based on the estimated RD relationship to improve the coding efficiency

and the presentation quality of a communication system [143]. Therefore, modeling of this RD relationship becomes a crucial issue in robust video codings and transmissions.

Previous proposed RD models, such as those in [144, 145], are often specified for predictive-coded video frame (P-frame) and hence cannot describe the $R(Q)$ and $D(Q)$ relationship for encoded sequence that involves intra-coded frame (I-frame). To solve this problem, models to describe the $R(Q)$ relationship separately for I and P-frames were suggested in [146]. Nevertheless, the approach in [146] is frame-based and it requires to switch between the estimated $R(Q)$ models, for different frame types, to obtain the overall $R(Q)$ relationship for a sequence of video frames. It requires high computational effort. Besides, the effects of motion estimation and compensation processes on the $R(Q)$ and $D(Q)$ relationships for I and P-frames have not yet been explored.

In this chapter, we shall briefly overview the existing RD estimation techniques for image and video compressions. In particular, the so-called ρ-domain analysis, which can uniquely characterize the RD relationship of an individual image or video frame for typical transform coding systems, will be discussed in detail. After that, we give a comprehensive analysis on the RD relationship for the compression of a sequence of video frames for block-based video codec. In particular, we study the relationship between the quantization process and the RD performance for block-based video codec with various GOP structures. We show that the average distortion of an encoded sequence is dominated by the quantization error of the I-frame involved. We also show that the $D(Q)$ function can be approximated by a linear model in terms of Q using regression analysis. Based on the proposed $D(Q)$ model, we further derive the $R(Q)$ model accordingly. The parameters of the proposed models are estimated using a least-square optimization approach. The estimated data are compared with the actual measurements and promising results are obtained.

5.2 Theory of RD Modeling

As mentioned in the previous section, RD theory seeks to characterize the relationship between coding rate and measured distortion in a consistent way. Classical RD analysis modeled the coding rate by considering the entropy of the quantized coefficients [137, 138]. However, they possess high degree of deviation between the practical and the theoretical estimations, especially in case of very low bit rates [147, 148, 149]. Consequently, many approaches were proposed to solve the above mismatch problem for the current transform coding systems and they can be categorized into two main classes [150]: analytical and empirical approaches.

Techniques using analytical approach often try to derive a set of mathematical formulas for describing the RD behaviors based on the statistical properties of the source data. In these techniques, the coding system and the image are first decomposed to be described by individual known statistical model and then combined to form a complete analytical model for the whole coding system. For example, in the analytical source model proposed in [141], a theoretical entropy formula for the quantized DCT coefficients was introduced based on the RD theory of the Gaussian source and the uniform quantizer. The error between the theoretical and practical values are then compensated by an empirical formula.

Empirical RD estimation techniques usually estimate the RD relationship by fitting the actual RD measurements to some empirical models. For example, in [151], eight actual RD points were suggested to construct a final RD curve using cubic interpolation. Similar empirical RD models have been proposed in [144, 145, 146]. In general, this type of RD estimation techniques requires to perform coding and decoding for different quantization parameters in order to obtain the actual RD measurements for model fitting which may lead to very high computational complexity.

However, existing RD models reported in the literature are quite different from each other for different coding algorithms. To solve this problem, a unified RD framework, called the ρ-domain analysis, was proposed in [147, 148, 149, 150] to approximate the RD relationships for most typical transform coding systems using a single model. In the following section, we shall briefly review the concept of ρ-domain analysis.

5.3 ρ-domain Analysis for RD Estimation

To uniquely characterize the RD relationship using a single approach for typical image and video transform coding schemes such as: EZW and JPEG coding, MPEG-2, H.263, and MPEG-4 video coding, the ρ-domain source modeling analysis was first introduced by He *et al* in [147].

Most transform coding algorithms often emphasize the efficient coding of zeros (which represent the irrelevant information of the source after quantization). There is a strong relationship between the percentage of zeros among the quantized transform coefficients and the RD behavior of the source encoder. In ρ-domain analysis, the percentage of zeros among the quantized coefficients is denoted by ρ. Different from the classical approaches which try to determine the RD relationship through the quantization parameter Q (or the so-called Q-domain) by exploring the $R(Q)$ and $D(Q)$ behaviors, the ρ-domain approach aims at mapping the Q-domain functions $R(Q)$ into the ρ-domain function $R(\rho)$, which has unique properties for typical image and video transform coding systems, as an alternative. That is, for a given value of Q, we can look for the corresponding value of ρ as:

$$\rho(Q) = \int_{-\Delta}^{+\Delta} f_D(x)dx \tag{5.1}$$

where Δ denotes the dead-zone threshold of the quantization scheme using the quantization step Q; and $f_D(.)$ refers to the pdf of the transform coefficients of the input image or motion compensated video frame, respectively.

Two characteristic rate curves, $Q_{nz}(\rho)$ and $Q_z(\rho)$, were introduced to characterize the non-zero and zero transform coefficients individually in the ρ-domain analysis as:

$$Q_{nz}(\rho) = \frac{1}{M} Q'_{nz} \tag{5.2}$$

and $$Q_z(\rho) = \frac{1}{M} Q'_z \tag{5.3}$$

where M corresponds to the size of the input image or video frame; Q'_{nz} and Q'_z denote the sum of the sizes of all non-zero coefficients and the sum of the sizes of all run-length numbers, respectively, after the quantization process in the pseudocoding scheme [148]. Based on experimental results, it was found that $Q_{nz}(\rho)$ is almost a straight line passing through the point (1,0). Hence, it can be modeled by the following linear expression:

$$Q_{nz}(\rho) = \theta(1-\rho) \tag{5.4}$$

where θ is a constant and can be obtained by:

$$\theta = \frac{Q_{nz}(Q_0)}{1-\rho(Q_0)}$$

where $Q_{nz}(Q_0)$ is the sum of the sizes of all non-zero quantized coefficients for any given quantization step Q_0, and $\rho(Q)$ is the one-to-one mapping function between the Q-domain and the ρ-domain, respectively. It was also observed that there is a strong correlation between θ and $Q_z(\rho_i)$ such that $Q_z(\rho_i)$ can be expressed in terms of θ using the following linear model:

$$Q_z(\rho_i) = A_i\theta + B_i \tag{5.5}$$

where A_i and B_i are the model parameters which can be obtained by statistical regression. It should be noted that, higher order polynomial can also be applied in Eq. (5.5) to achieve

better estimation accuracy [148]. Finally, the rate curve decomposition scheme $R(\rho)$ is approximated by $\hat{R}(\rho)$ which is a linear combination of $Q_{nz}(\rho)$ and $Q_z(\rho)$ as:

$$\hat{R}(\rho) = \xi_1(\rho)Q_{nz}(\rho) + \xi_2(\rho)Q_z(\rho) + \xi_3(\rho) \tag{5.6}$$

where $\xi_i(\rho)$ are the model parameters. By using the Q-ρ mapping function, $R(\rho)$ is then mapped into the Q-domain to obtain $R(Q)$.

Based on Eqs. (5.4)-(5.6), a unified RD framework which has unique properties for all typical transform coding systems is provided. This RD framework can be applied for frame-based rate controls in image and video codings [147, 148, 149, 150] as well as mode-selection for error resilient video codings [64]. However, the proposed ρ-domain RD framework works only for an individual image or video frame. It cannot be used to describe the RD relationship for a sequence of video frames.

In fact, the RD estimation in coding a sequence of video frames can help to perform macroscopic analysis and improve the efficiency of the transform coding schemes. For example, when coding a sequence of video frames, if we know the RD relationship between the first I-frame and any P-frame within a GOP, we can perform better bit allocation between the I-frame and the P-frames in the GOP with lower computational complexity than the frame-based approaches.

In the subsequent sections, we shall give a thorough analysis for the relationship between the quantization process and the RD performance for a sequence of video frames encoded with various GOP structures using a block-based video codec.

5.4 GOP-based RD Analysis

5.4.1 Final distortion vs distortion from I and P-frames

Without losing the generality, let us consider the video sequences that are encoded by a block-based H.263 [11] video coder, which involves I and P-frames only, to elaborate the idea of GOP-based RD analysis. The basic configuration of the H.263 video source coding algorithm is based on a hybrid of inter-picture prediction to reduce temporal redundancy and transform coding of the residual signal to reduce spatial redundancy. By using this algorithm, I-frames are encoded independently and P-frames are predicted and encoded with respect to the previous frame. To simplify our discussion, other frame types, for example, the bi-directional-encoded frame (B-frame), are not considered in this chapter.

In order to study the relationship between the quantization process and the RD performance, we measure the distortion as the MSE between the original uncoded sequence and the decoded sequence frame by frame. The distortions of luminance ($D_{i,Y}$) and chorminance ($D_{i,U}$ and $D_{i,V}$) components for each frame are formulated as:

$$D_{i,j} = \frac{1}{NM} \sum_{x=0}^{N-1} \sum_{y=0}^{M-1} \left| f_{i,j}(x,y) - \hat{f}_{i,j}(x,y) \right|^2 \tag{5.7}$$

for $i = 1,2,\ldots,k$; $j = Y$, U and V components

where $|x|$ denotes the absolute value of x; i is the frame index; k is the sequence length; N and M correspond to the frame width and height; $f_{i,j}(x,y)$ and $\hat{f}_{i,j}(x,y)$ are the pixel value for the j-component at the position (x,y) of the original and the decoded ith-frame, respectively. From Eq. (5.7), the average distortion, D_R, for k decoded frame at coding rate R is computed as:

$$D_R = \frac{1}{3k} \sum_{i=0}^{k-1} \left[D_{i,Y} + D_{i,U} + D_{i,V} \right] \tag{5.8}$$

where R is the average bit rate (per second) used to encode the k frame. In fact, the distortion, as described in Eq. (5.8), is mainly introduced from the motion estimation and the quantization processes. However, in case of error-free transmission, the distortion coming from the motion estimation process is compensated by encoding and transmitting the residual error in the prediction algorithm and hence it is negligible to the final distortion. Consequently, the lossy quantization process among the DCT coefficients is the main source of the final distortion, D_R, in Eq. (5.8). Actually, the MSE in pixel-domain is the same as that in the DCT-domain [152]. Nevertheless, D_R cannot be computed directly by finding the difference between the dequantized and the original DCT coefficients. It is because the distortion due to the quantization among the DCT coefficients in different frame types gives different levels of contribution to D_R.

It can be shown that D_R is dominated by the quantization error from the DCT coefficients in I-frame while that in P-frame is insigificant. This phenomenon can be explained by exploring the coding natures of I and P-frames. Indeed, the content in P-frame is the residual difference between the current frame and the previous encoded frame after prediction. Error introduced by quantizing the DCT coefficients of this residual error gives only minor effect to the final distortion of the decoded sequence, especially if the motion of the video sequence being encoded is gentle. Furthermore, in H.263 video coding, DC and AC DCT coefficients are quantized separately in each transformed block [11]. Different from the quantization of AC coefficients using a midrise quantizer with step size equals to $2Q$, all DC coefficients are mapped to a 8-bit codeword to improve their accuracy. As a result, the current predicted frame can still be reconstructed roughly by referring to the content of the previous frame, the corresponding motion vectors, and the DC coefficients in each block without the AC coefficients. Thus, the information loss due to the quantization of DCT coefficients in P-frame does not contribute significantly to the final distortion in Eq. (5.8). Nevertheless, due to the independent coding strategy for the content of I-frame,

quantization error of DCT coefficients in I-frame gives direct effect to the distortion in Eq. (5.8). In H.263 video coding, the relationship between an I-frame and its subsequent P-frame can be modeled using the following equation:

$$\begin{aligned}\varepsilon_P &= M(P,\hat{I}) + MV_P \\ &= M(P, I\text{-}E_I^{Q_I}) + MV_P \\ &= M(P,I) + MV_P + E_I^{Q_I}\end{aligned} \tag{5.9}$$

where ε_P denotes the residual error after the motion estimation; $M(.)$ represents the motion estimation operator; I and P denote the content of I and P-frames, $\hat{I}$ denotes the reconstruction of the I-frame; MV_P and $E_I^{Q_I}$ represents the corresponding motion vectors and the quantization error in I-frame using quantization step Q_I, respectively. From Eq. (5.9), based on the assumption that the frame difference and the quantization errors in the previous frame are uncorrelated, if the motion vector can be transmitted without any loss, the quantization error (measured in terms of MSE) in the P-frame using the quantization step Q can be simplified as:

$$\begin{aligned}D\{\varepsilon_P\} &= D\{M(P,I)\} + D\{E_I^{Q_I}\} \\ &= \int_{-2.5Q}^{2.5Q} x^2 f_M^{P,I}(x)dx + \int_{-2.5Q}^{2.5Q} x^2 g_E^{I,Q_I}(x)dx\end{aligned} \tag{5.10}$$

where $D\{.\}$ denotes the distortion measurement; $f_M^{P,I}(.)$ stands for the pdf of the transform coefficients after motion estimation between the I and P-frames; $g_E^{I,Q_I}(.)$ represents the pdf of quantization error in I-frame using quantization step Q_I, and Q denotes the quantization step of P-frame, respectively. It should be noted that in H.263 video coding, all intra-coded transform coefficients having amplitudes smaller than $2Q_I$ will be quantized to zeros. However, in the inter-coded mode, only those transform coefficients having amplitudes smaller than $2.5Q$ will be considered as zeros. Therefore, we know that:

$$Max\{g_E^{I,Q_I}\} < 2Q_I$$

and hence, if

$$2.5Q \geq 2Q_I \tag{5.11}$$

then, from Eq. (5.10),

$$\begin{aligned} D_P = DP\{\varepsilon_P\} &= \int_{-2.5Q_P}^{2.5Q_P} x^2 f_M^{P,I}(x)dx + D_I \\ &\geq D_I \end{aligned} \tag{5.12}$$

where D_I and D_P represent the final distortion measurements in the I-frame and the P-frame, respectively. It can be shown from Eq. (5.12) that D_P has value always greater than D_I. Furthermore, if the content of these two frames are similar (this happens particularly for the frames within the same GOP which share the same scene), Eq. (5.12) can be further simplified as:

$$D_P \approx D_I . \tag{5.13}$$

Eqs. (5.12) and (5.13) can be generalized to the other P-frames in the same GOP by carefully replacing the reconstruction of their previous encoded frame. Eq. (5.13) indicates that we can approximate the distortion of P-frames by referring to the distortion of I-frame in the same GOP if Eq. (5.11) holds. Indeed, most video coding strategies employ larger quantization step values for coding P-frame than that for I-frame to achieve the goal of compression. Hence, Eq. (5.11) is valid for general video transform codings.

To further elaborate the above explanation, let us take a look at some real examples. In our examples, five standard QCIF video sequences, Car-phone, Claire, Foreman, Mother and daughter, and Salesman, are encoded and decoded at 30 frames/sec by using a H.263 codec with all advanced feature disabled. In order to investigate the effects of quantization errors of I and P-frames on the final distortion, the video sequences are encoded at three different sequence structures: (i) 1 I-frame only; (ii) 1 I-frame with 19 P-frames, and (iii) 1 I-frame with 99 P-frames. All frames are quantized by the same Q with the quantizer suggested in H.263. For each sequence structure, various values of Q are applied and the distortions between the uncoded sequences and the decoded sequences are measured

according to Eqs. (5.7) and (5.8). Then, the final distortion values are converted to PSNRs for comparison. Table 5.1 shows the resulted PSNR versus Q of the test sequences for the defined sequence structures, respectively.

It is seen from Table. 5.1 that the distortions of coding structure (i) follow closely to those of structure (ii) and (iii) for all test sequences. Besides, the average difference in the PSNR measurements between the coding structure (i) and (iii) is only 0.69 dB. These small differences in the overall distortion are due to the quantization error on the residual information left after the motion prediction. It is also shown from Table 5.1 that when Q becomes larger, the overall distortions of the three structures are closed to each other. These results show that the final PSNR measurements are dominated by the distortions from the I-frame when most AC coefficients in the P-frames are quantized to zero as described in Eq. (5.13). On the other hand, it is shown from Table 5.1 that the PSNR measurements of structure (ii) and (iii) are very close to each other all the test sequences although structure (iii) has 80 more P-frames than structure (ii). In fact, the average PSNR differences between the coding structure (ii) and (iii) is 0.17 dB only. It means the contribution to the final distortion from the P-frames is diminutive.

As a brief summary, the final distortion of the decoded sequence, with respect to the original uncoded sequence, can be estimated by merely considering the distortion due to the quantization of I-frame while that of P-frame is negligible.

		Average Distortion, D_R, in PSNR (dB)			
Test sequence	Coding structure	Quantization value (Q)			
		5	11	19	26
Car-phone	1-I	40.50	36.20	33.68	31.87
	1-I, 19-P	39.68	35.68	33.43	31.97
	1-I, 99-P	39.65	35.48	33.12	31.64
Claire	1-I	42.04	37.51	35.18	33.50
	1-I, 19-P	41.15	36.98	34.81	33.20
	1-I, 99-P	41.09	36.92	34.74	33.29
Foreman	1-I	39.58	35.49	33.16	31.44
	1-I, 19-P	38.44	34.59	32.46	31.01
	1-I, 99-P	38.25	34.24	32.01	30.48
Mother and daughter	1-I	39.81	35.74	33.67	32.19
	1-I, 19-P	39.16	35.41	33.52	32.27
	1-I, 99-P	39.01	35.14	33.15	31.87
Salesman	1-I	39.12	34.78	32.49	31.03
	1-I, 19-P	38.64	34.23	32.21	30.86
	1-I, 99-P	38.59	34.35	32.11	30.72

Table 5.1: List of D_R values for five QCIF sequences encoded at three GOP structures with different values of quantization value Q.

5.4.2 Proposed GOP-based models

From the previous section, it is known that the distortion of I-frame can be used to predict the overall distortion of a GOP. Indeed, the distortion of I-frame is introduced from the quantization error of its DCT coefficients. Nevertheless, as mentioned before, all DC DCT coefficients are encoded accurately by mapping them into a 8-bit codeword. Therefore, quantization errors among the AC coefficients are the major source for the distortion in I-frame. In this section, we try to derive the $D(Q)$ and $R(Q)$ models based on the simplest GOP structure that involves only 1 I-frame and finite number of P-frames and then extend them to other possible GOP structures.

In practice, all AC coefficients in I-frame are quantized by an uniform midrise quantizer with dead-zone around zero as:

$$I(i,j) = \left\lfloor \frac{|x(i,j)|}{2Q} \right\rfloor \tag{5.14}$$

where $\lfloor k \rfloor$ is the rounding operator denoting the integral part of the real number k; $I(i,j)$ and $x(i,j)$ are the quantization index and the original AC coefficient at position (i,j), respectively. The reconstructed coefficient can be computed as:

$$\hat{x}(i,j) = 2Q \cdot I(i,j)\,. \tag{5.15}$$

If we express $|x(i,j)|$ as:

$$|x(i,j)| = 2k(i,j) \cdot Q + \zeta(i,j)$$

where $k \in (Z \geq 0)$ and $\zeta \in (0, 2Q]$, then

$$I(i,j) = \left\lfloor \frac{2k(i,j) \cdot Q + \zeta(i,j)}{2Q} \right\rfloor = \begin{cases} 0, & \text{if} \quad k(i,j) = 0 \\ k(i,j), & \text{if} \quad k(i,j) > 0 \end{cases}$$

and $$\hat{x}(i,j) = \begin{cases} 0, & \text{if} \quad k(i,j) = 0 \\ 2Q \cdot k(i,j), & \text{if} \quad k(i,j) > 0\,. \end{cases}$$

The distortion for the J-colour component in an I-frame, can be calculated as:

$$D_I^J(i,j) = \begin{cases} x^2(i,j), & \text{if} \quad k(i,j) = 0 \\ \xi^2(i,j), & \text{if} \quad k(i,j) > 0. \end{cases}$$

From the equations above, the distortion, D_R, in Eq. (5.8) can be computed as:

$$D_R(Q) \approx D_I(Q) = \int_{-2Q}^{2Q} x^2 f_I(x)dx + \int_{-2Q}^{2Q} \xi^2 g_I(\zeta)d\zeta \left\{ \int_{2Q}^{\infty} f_I(x)dx + \int_{-\infty}^{-2Q} f_I(x)dx \right\} \tag{5.16}$$

where $f_I(x)$ and $g_I(\zeta)$ are the pdf of the AC coefficients and the remaining coefficients after the modulus of the AC coefficients for all colour components in the I-frame, respectively. However, due to the fact that the DCT process shifts signal energy to lower frequency coefficients, and most AC coefficients are quantized to zero when a large value of Q is applied. Eq. (5.16) can be simplified as:

$$D_R(Q) \approx \int_{-2Q}^{2Q} x^2 f_I(x)dx. \tag{5.17}$$

From Eq. (5.17), we know that D_R can be calculated by considering the pdf of the AC coefficients of an I-frame. Indeed, the pdf of the AC coefficients for a transformed block (usually 8×8 pixels) are often considered as Laplacian [153]. However, in practice, certain extent of deviation can be found that can significantly affect the RD estimation of a video frame. We suggest to use the following model instead to describe the pdf of the AC coefficients for an I-frame:

$$f(x) = \frac{K}{N_x} e^{-(|x|/\alpha)^{\beta}} \tag{5.18}$$

where K represents the proportional constant; α and β are arbitrary positive constants; N_x denotes the total number of AC coefficients in the frame. Fig. 5.1 compares the actual pdfs of the AC coefficients of the I-frames to the best-fitted Laplacian model and the suggested model in Eq. (5.18) for two standard video sequences (in QCIF format), Claire and Foreman, respectively.

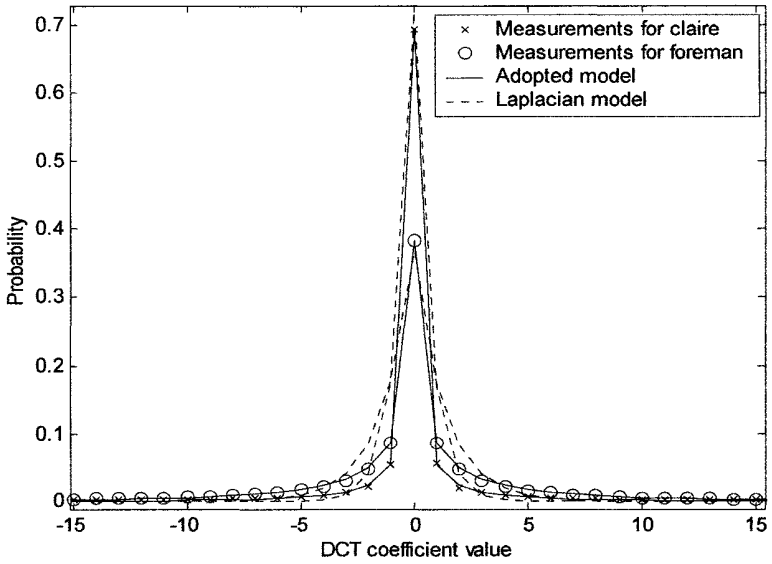


Figure 5.1: The "zoom-in" pdfs of the DCT coefficients for the I-frame of the two test sequences and their best-fitted Laplacian model as well as the adopted model in Eq. (5.18).

It is shown clearly from Fig. 5.1 that the pdfs of the AC coefficients, of the I-frames for both Claire and Foreman sequences, are deviated from the Laplacian model but well approximated by the suggested model. Two other standard sequences, Car-phone and Salesman are also tested and similar results are obtained. These results show that it is safe for us to model the pdf of the AC coefficients of an I-frame for most standard sequences using Eq. (5.18).

In H.263 standard, the sign of DCT coefficient is neglected during the quantization process. By combining Eqs. (5.17) and (5.18), Eq. (5.17) can be expressed as:

$$D_R(Q) = \frac{2K}{N_x} \int_0^{2Q} x^2 e^{-(x/\alpha)^\beta} dx. \tag{5.19}$$

However, it can be proven that the above integral involves an incomplete gamma function that cannot be expressed explicitly in closed form using familiar functions and constants of calculus. Fortunately, the variable Q in Eq. (5.19) is a controlled variable that can be measured without substantial error. Thus, the dependence of D_R on Q can be estimated by

regression analysis. To ease our discussion, we discretize the DCT coefficients before performing the quantization. It should be noted that this discretization process would only give negligible error on the calculation of D_R without losing the generality. Consequently, Eq. (5.19) can be rewritten as:

$$D_R(Q) = \frac{2K}{N_x} \sum_{x=0}^{2Q-1} x^2 e^{-(x/\alpha)^\beta} \quad . \tag{5.20}$$

Hence, $D_{i,j}$ can be obtained by estimating the arbitrary constant K, α and β from Eq. (5.18) with respect to the pdf of the AC coefficients of an I-frame for a particular sequence. The dependence of D_R on Q can be found by examining the correlation coefficient (μ) defined as [154]:

$$\mu = \frac{n\sum_{k=0}^{n-1} Q_k D_R(Q_k) - \sum_{k=0}^{n-1} Q_k \sum_{k=0}^{n-1} D_R(Q_k)}{\sqrt{n\sum_{k=0}^{n-1} Q_k^{\,2} - \left(\sum_{k=0}^{n-1} Q_k\right)^2} \cdot \sqrt{n\sum_{k=0}^{n-1} D_R^{\,2}(Q_k) - \left(\sum_{k=0}^{n-1} D_R(Q_k)\right)^2}} \tag{5.21}$$

where n is the total number of measurements. Table 5.2 shows the estimated pdf model parameters for the AC coefficients in the I-frames and the corresponding values of μ, of five QCIF sequences with $Q \in [0,31]$.

	Model parameters			Correlation coefficient (μ)
Test sequence	K	α	β	
Car-phone	0.4685	0.1098	0.3266	0.9990
Claire	0.6872	0.0401	0.2917	0.9990
Foreman	0.3838	0.1990	0.3551	0.9990
Mother and daughter	0.3740	0.3263	0.3905	0.9981
Salesman	0.4537	0.1921	0.3679	0.9973

Table 5.2: List of estimated model parameters to describe the pdfs of transform coefficients of the I-frames using Eq. (5.18) for five QCIF sequences and the corresponding correlation coefficients in their $D(Q)$ measurements.

From Table 5.2, the values of μ for all sequences are over 0.99 which indicate that their $D(Q)$ functions are perfectly linear, by regression analysis, such that we can write:

$$D_R \propto Q$$

and so, D_R in Eq. (5.8) can be rewritten as:

$$D_R(Q) = aQ + b \tag{5.22}$$

where a and b are arbitrary constant. For other possible GOP structures, encoded sequence can be regarded as the combination of several independent GOP segments involving 1 I-frame and a finite number of P-frames that results in a linear model as follows:

$$\begin{aligned} D_R(Q) &= \frac{1}{\tau}\sum_{\tau=0}^{\tau-1}[a_i Q + b_i] \\ &= AQ + B \end{aligned} \tag{5.23}$$

where τ represents the number of segments. For the $R(Q)$ function, it can be shown from [155] that for any given source with variance δ^2, there exists a block code of fixed rate R as:

$$R > \frac{1}{2}\ln\left(\frac{\delta^2}{D}\right) \tag{5.24}$$

where D is the average distortion. Indeed, for many sources and distortion measures, the lower bound of the RD relationship coincides with the actual RD function for some fidelity criterion on the distortion measurement at lower range [155]. Therefore, based on the above theory, the RD relationship can be formulated as:

$$R(D_R) = \frac{1}{2}\ln\left(\frac{\delta^2}{D_R}\right) \tag{5.25}$$

where δ^2 is the variance of the DCT coefficients and D_R is the average distortion defined in Eq. (5.8). By combining Eqs. (5.23) and (5.25), the $R(Q)$ relationship can be expressed as:

$$R(Q) = \frac{1}{2}\ln\left(\frac{\delta^2}{aQ+b}\right).$$

Similar to [146], the above expression can be expanded into Taylor series as:

$$R(Q) = c + dQ^{-1} + eQ^{-2} + R_3(Q) \tag{5.26}$$

where c, d, and e are the polynomial coefficients and $R_3(.)$ is the tail of the series at order-3, respectively. Eq. (5.26) can be further simplified as:

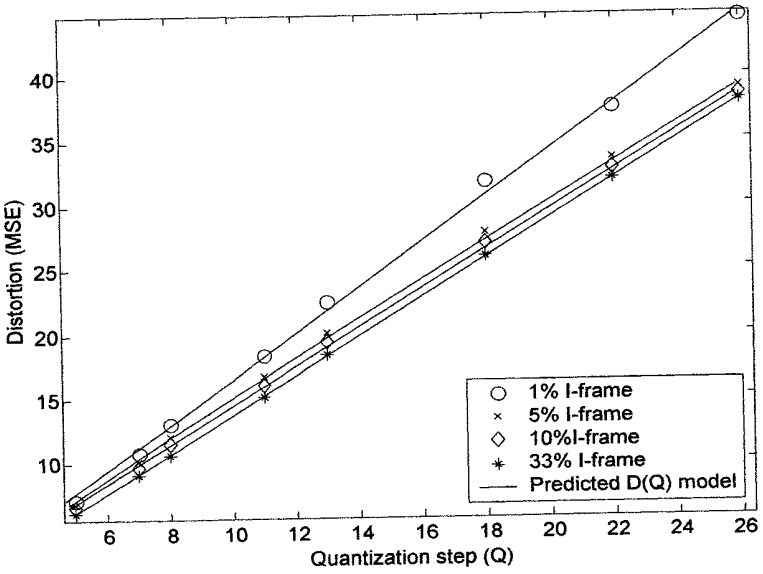
$$R(Q) = CQ^{-1} + DQ^{-2} + E \tag{5.27}$$

where C, D, and E are arbitrary constants. It should be noted that Eq. (5.26) is valid for all range of GOP structure as it is derived from Eq. (5.23). The proposed $R(Q)$ model in Eq. (5.27) is similar to the one in [146] but it describes the average RD performance for the whole encoded sequence instead of an individual frame. It should be noted that all parameters in Eqs. (5.22), (5.23), and (5.27) are sequence and GOP structure dependent.

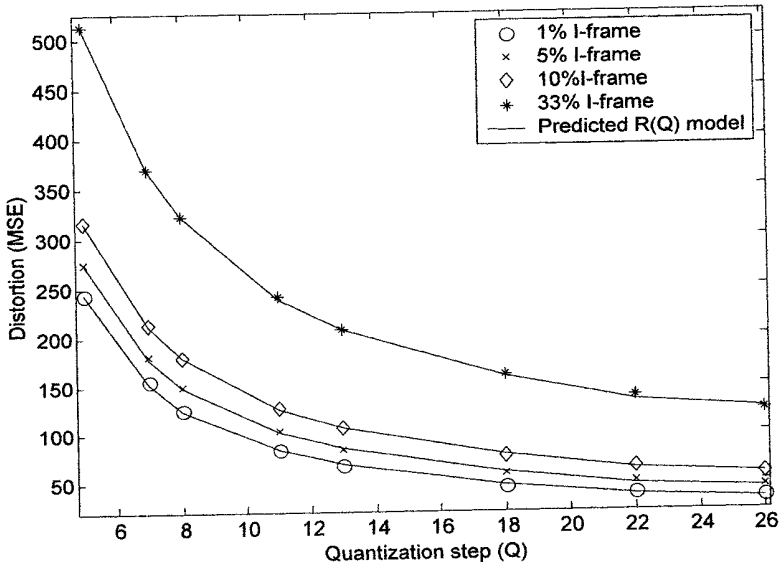
5.4.3 Simulation results and discussions

Systematic study is carried out to investigate the accuracy of the proposed $D(Q)$ and $R(Q)$ models. Five standard QCIF sequences are tested by using a baseline H.263 codec with all

advanced features disabled. In the simulation, 100 frames are encoded at 30 frame/sec with different percentages of I-frame (γ) using various Q to achieve different R for each sequence. The $D(Q)$ and $R(Q)$ curves for different γ are obtained by interpolating the actual measurements. The model parameters in Eqs. (5.23) and (5.25) are calculated by minimizing the MSE between four actual measurements and the models using a least-square approach. All distortion measurements are then converted to the PSNR for comparison. Figs. 5.2 and 5.3, show the predicted $D(Q)$ and $R(Q)$ values, together with eight actual measurements, for the Car-phone and Claire sequences at different γ, respectively. The average biases (absolute difference) between the actual measurements and the estimated $D(Q)$ and $R(Q)$ models for all the test sequences are listed in Table 5.3.

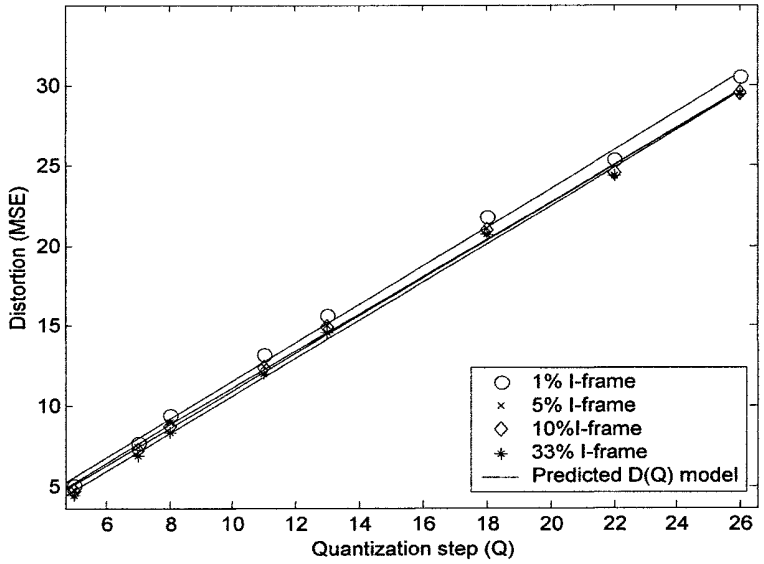


(a)

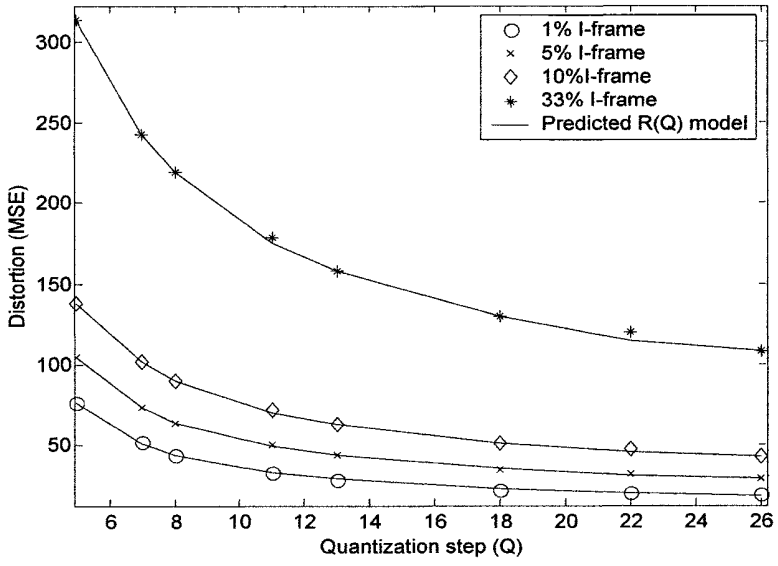


(b)

Figure 5.2: (a) $D(Q)$ and (b) $R(Q)$ curves for the sequence Car-phone encoded with γ = 1%, 5%, 10% and 33%.



(a)



(b)

Figure 5.3: (a) $D(Q)$ and (b) $R(Q)$ curves for the sequence Claire encoded with γ = 1%, 5%, 10% and 33%.

Test sequences	Average bias	
	Distortion (PSNR in dB)	Rate (Kbits/sec.)
Car-phone	0.184	1.144
Claire	0.219	0.476
Foreman	0.226	1.182
Mother and daughter	0.172	0.555
Salesman	0.170	2.086

Table 5.3: The average biases for the estimated models to the actual measurements.

From Figs. 5.2 and 5.3, it is observed that the predicted values follow the actual measurements closely for both test sequences. And from Table 5.3, the average bias for the estimated distortions and rates among the five sequences are only 0.194 dB and 1.089 Kbits/sec. These results provide solid support for the accuracy of the proposed models. The proposed models use only five parameters and thus requires reasonable complexity for optimization. By using the proposed models, fast GOP-based rate control can be done easily by adjusting Q to achieve specific coding rate and distortion with much lower computational effort than the frame-based one. In fact, we shall show in the next chapter that the proposed GOB-based models can be used to simplify the optimization process in coding mode selection for error resilient video coding.

5.5 Summary

In this chapter, we have given an overview of the current RD estimation techniques for image and video compressions. In particular, a unified RD analysis framework called the ρ-

domain analysis, which uniquely characterizes the RD relationship for most typical transform coding systems, has been reviewed. Nevertheless, the ρ-domain analysis can only provide accurate RD estimation for an individual image or video frame. Hence, if a GOP of a video sequence is considered, we have to carry out RD estimation for each frame and it requires intensive computational complexity. To solve the problem, we give a comprehensive analysis on the RD relationship of a GOP for block-based video codec. We show that the average distortion of a GOP is dominated by the quantization error among the I-frame in the GOP and the $D(Q)$ function can be approximated by a linear model in terms of Q using regression analysis. Based on the proposed $D(Q)$ model, we derive the $R(Q)$ model accordingly. The parameters of the proposed models are estimated using a least-square optimization approach. The estimated data are compared with the actual measurements and promising results are obtained.

Chapter 6 Optimally Using Mode-selection and Resynchronization Marker for Error Resilient Video Coding

6.1 Introduction

Recent advances in wireless communication technology have provided a much wider bandwidth for video transmission. However, channel errors due to different problems in wireless communications, such as fading, often disturb the compressed images and videos and consequently degrade the performance of most multimedia communication systems. Reliable transmission of compressed video over error-prone channel therefore becomes an important research topic. Previous error resilient video communication techniques can be classified into two categories: with and without feedback channel. If feedback channel is available, error can be recovered accurately by adopting automatic repeat request (ARQ) or error tracking techniques [72]. However, providing feedback channel can be expensive in certain applications. Besides, it can be also affected by the round-trip delay and channel errors of the network. In this chapter, we focus on those techniques that try to strengthen the bitstream against channel errors for robust transmission without feedback channel. Mode-selection [99, 100] for intra-refreshing and resynchronization marker (RM) [71, 156] are two commonly used approaches in this class as mentioned in Section 2.4.4.

Decoder loses synchronization with encoder when decoding a damaged video bitstream. It causes spatial and temporal error propagation in the reconstructed frames until

correct intra-coded information (either intra-coded frames or macroblocks) is received. The objective of the mode-selection approaches is to selectively intra-code some video frames or macroblocks (MBs) to stop error propagation. However, it reduces significantly the coding efficiency, as compared with the inter-coded case.

The use of RM is another effective way to restrict error propagation. Predefined markers are introduced to the video bitstream to allow the decoder to regain synchronization in case of error. However, frequent insertions of RM will also introduce a huge overhead to the bitstream and degrade the compression efficiency. Hence, the positioning of RM has been studied extensively [157, 158]. For a given number of RMs, these approaches try to estimate their positions such that optimum performance in RD sense can be achieved. The approaches require the number of RMs as an apriori information which is very much dependent on the channel conditions and video contents. In fact, more RMs are needed if the channel error rate is high or the video has vivid motion. The relationships between the number of RMs, channel error rate, and video contents have not been addressed quantitatively in these approaches and their determinations are difficult in practice.

Although the RM approaches allow resynchronization, error still remains in the video and persists for a long time if it is not intra-refreshed. Intra-coding a frame/MB by mode-selection helps to refresh the damaged contents but it causes more overhead than the RM approaches. To achieve accurate error detection and recovery at a reasonable level of overhead, both the mode-selection and the RM approaches should be applied. However, a mechanism is needed to determine when the RM approaches or the mode-selection should be applied. The mechanism must have low computational complexity or otherwise it will be difficult for practical implementation.

In this chapter, we present an optimization scheme for error resilient video coding. A new RM, called the macroblock RM (MBRM), is proposed. Noted that we do not fix the number of MBRMs to be used for a frame, but propose a RD-based mode-selection

algorithm to optimize the coding mode selection and the use of MBRM in the bitstream. The algorithm firstly estimates the overall distortion for each group-of-block (GOB) of a decoded frame by considering the effects of quantization, error propagation, and error concealment. The best coding mode of each GOB is then determined to give the minimum estimated distortion for a given rate budget. We verify the proposed scheme by applying it to a one-way video communication system. Simulation results show that, for general video conferencing or surveillance applications, the proposed scheme gives significant gain up to 8 dB in PSNR in the decoded videos as compared with the traditional approaches for error-prone channel with BER ranged from 10^{-5} to 10^{-3}.

This chapter is organized as follows. In the next section, a brief review on the applications of mode-selection and RM in error resilient video coding is provided. We then introduce a novel approach by jointly optimizing the uses of the mode-selection and the RM for error resilient video coding in Section 6.3. The proposed MBRM and the GOB-based mode-selection algorithm, as well as the potential application of the proposed scheme are discussed. Simulation settings and results of the proposed scheme, comparing with the traditional techniques, are given in Section 6.4. Finally, we summarize this chapter in Section 6.5.

6.2 Use of Mode-Selection and Resynchronization Marker for Error Resilient Video Coding

6.2.1 Mode-selection

As mentioned in the previous section, channel errors often destroy the transmitted bitstream. The error in the corrupted bitstream may further propagate both spatially and temporally in the reconstructed frames due to the extensive uses of motion estimation and compensation in most transform video coding schemes until correct intra-coded information is received. If more intra-coded information is provided, the probabilities of spatial and temporal error propagation will be lowered. In the best case, temporal error propagation can be avoided if all the video frames being transmitted are intra-coded. However, intra-refreshing a video frame significantly decreases the coding efficiency of the encoder. Besides, channel errors can also corrupt the intra-coded frames and cause spatial error propagation among MBs. A more appropriate idea is to perform intra-refreshing at MB level rather than at the frame level. The question here is to determine when and which MB needs to be intra-refreshed in coding and transmitting a video sequence. Most mode-selection schemes, as discussed in Section 2.4.4, try to solve this problem by adopting different strategies to decide the coding mode of each MB in a video frame such that the propagation of error is limited.

Obviously, if feedback channel between the encoder and the decoder is available, the above problem can be solved easily by sending the information of the damaged MBs from the decoder to the encoder. If no feedback channel is provided, encoder has to determine the solution by itself or use information provided by the network channel. In general, the number of intra-refreshed MBs highly depends on the channel quality and capacity as well as the video content. If the channel condition is poor and the video sequence involves vivid motion, more MBs may need to be intra-refreshed and vice verse. Besides, we also need to consider the effect of error propagation and error concealment

adopted in the decoder when determining the coding mode of an MB. It is because the information carried by each MB is of different levels of importance. The corruption of the MBs, which contain moving objects or are used comprehensively for prediction and motion estimation/compensation, may introduce more severe distortion to the reconstructed frames than those MBs involve only static background. In fact, MBs located in low activity region can be recovered effectively by performing appropriate error concealment apart from using the expensive intra-refreshing scheme. The choice of MB for intra-refreshing is crucial to the final quality of the reconstructed frames.

6.2.2 Resynchronization marker (RM)

Resynchronization marker (RM) appears in the form of unique bit pattern in a bitstream such that it can be distinguished in the syntax level. RM helps decoders to regain synchronization in case of error and to reduce spatial error propagation. Besides, it is an effective tool for error isolation in a corrupted bitstream. The basic operation of RM has been discussed briefly in Section 2.4.4. In this section, we further study the use of RM in block-based video coding and transmission.

Without losing the generality, we shall use H.263 as the platform to illustrate the application of RM for bitstream resynchronization in case of error. In H.263, the fixed RMs (FRM), called PSC and GBSC, are located at each starting positions of the picture and GOB layers [12]. The working principle of this FRM approach is illustrated in Fig. 6.1. As shown in Fig. 6.1, when error occurs, all suspected data starting from the error detected position to the next resynchronization point will not be decoded correctly. Due to the difference between the error detected position and the actual error position, residual error is resulted. If only the suspected data are discarded, the residual error will propagate among the decoded frames and cause a serious corruption to the reconstructed video frames.

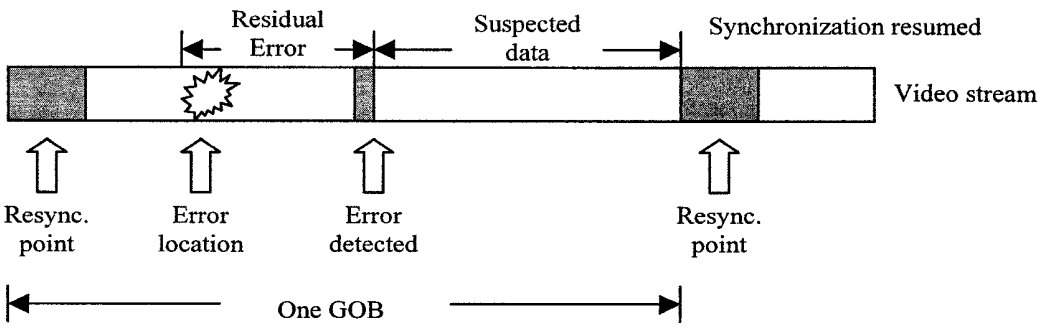


Figure 6.1: Working principle of resynchronization marker in GOB.

One solution to the problem is to discard all the data between the previous RM and the next valid RM when error is found. In H.263, at least one GOB/slice will need to be discarded when error occurs which may cause a serious information loss to the decoded frames. Later version of H.263 (H.263+) and MPEG-4 improve the situation by introducing the concept of video packet, which corresponds to an integer number of consecutive MBs. However, the effect will still be noticeable if a video packet is discarded due to the residual error. It should be noted from Fig. 6.1 that synchronization can only be resumed when a valid RM is received while uncorrupted MBs, if they locate at the damaged region, still need to be discarded no matter they are intra or inter-coded. Hence, the use of mode-selection in combination with RM is a critical condition in error resilient video coding.

6.3 Jointly Optimize the Use of Mode-Selection and Resynchronization Marker for Error Resilient Video Coding

6.3.1 Use of resynchronization marker in macroblock level for robust transmission

To stop error propagating from a corrupted MB to an error-free MB, it is desirable to introduce RMs as many as possible for error isolation. As the basic unit for a compressed video is block, it is intuitively to consider if RM can be applied among blocks. However, it is impossible to insert RM into the block layer for H.263 due to its irregular code pattern. Besides, if resynchronization is done among blocks, huge amount of RM is needed. Thus, we consider next to insert RM into the MB layer.

In this section, a new RM namely, the macroblock RM (MBRM), is proposed. The proposed MBRM, as shown in Fig. 6.2, is separated into two fields: the resynchronization word (12 bits) and the MB number indicator (4 or 5 bits depends on QCIF or CIF format). The MBRM is inserted to each MB except the skipped MBs since, first, the probability of error occurred in skipped MBs is very low; second, they can be exactly recovered by replacing it using the same MB in the previous frame.

Resynchronization word (12 bits) 0000 0000 0001	MB number (4 or 5 bits)

Figure 6.2: Structure for the proposed MBRM.

i. Operation principle

Fig. 6.3(a) illustrates the operation principle of the proposed MBRM. During decoding, decoder checks for the GBSC and MBRM and marks the previous decoded MBs that have MBRM. When error occurs, it searches for the next valid MBRM/GBSC for resynchronization. A valid MBRM should satisfy two criteria:

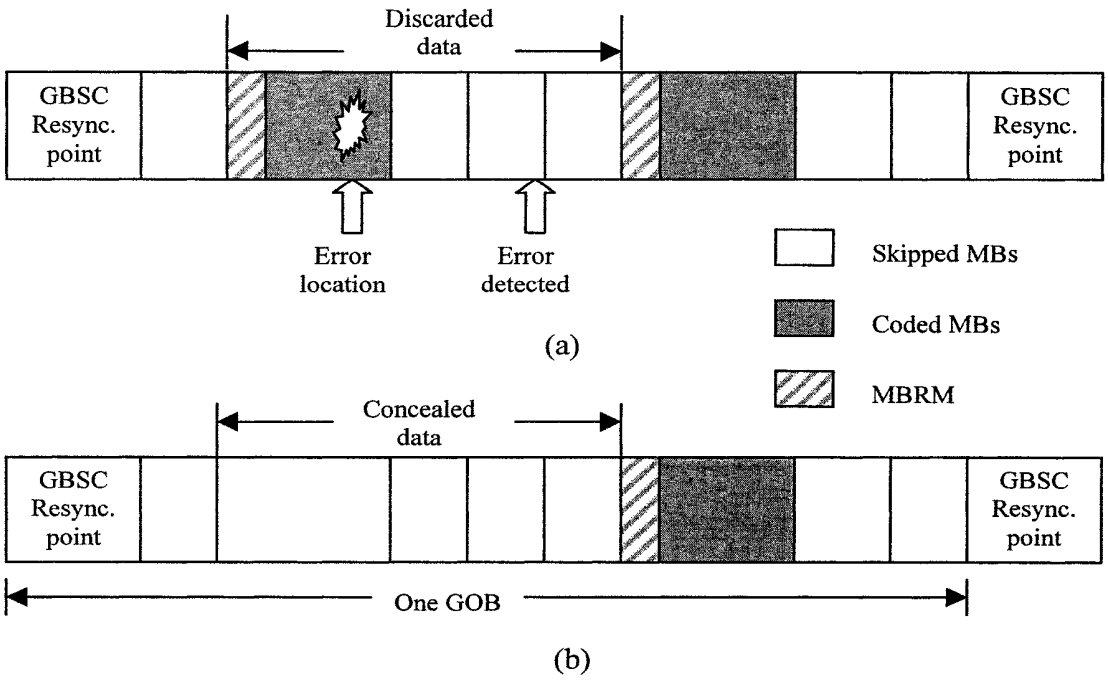


Figure 6.3: (a) Operation process for decoder using MBRM in case of error; (b) recovered bitstream after concealment.

1. A valid resynchronization marker, i.e. 11 zeros followed by a one.
2. The current MB number should be larger than that of the previous marked MB but not excess the maximum MB number in a GOB.

After locating the next valid RM, decoder discards the data between the previous marked MB and the next valid RM followed by an error concealment. In our case, we set the motion vectors of the discarded MBs to zero and copy the corresponding MBs from the previous frame to the corrupted frame for concealment. In H.263, it can be done by setting the discarded MBs to the type of skipped MBs as illustrated in Fig. 6.3(b). As shown in Fig. 6.3(a), in some cases, error propagates from corrupted MB to other error-free MBs, particularly the skipped MBs. It results in a false counting of the number of MBs since there is no information to show how many MBs need to be discarded. That is why the MB

number indicator is included in the MBRM. By comparing the MB number indicators in the previous and the next MBRM, the number of discarded MBs can be regained and the residual error can be eliminated. It is shown in Fig. 6.3(b) that even no MBRM is attached to the skipped MB, the skipped MBs can still be recovered correctly after concealment. As compared with the traditional RM approaches, the proposed approach localizes error effectively so that accurate concealment and resynchronization can be performed.

ii. Overhead consideration

It is interesting to note that, while the objective of video coding is to reduce the redundancy in the bitstream as much as possible, many error resilient coding techniques often try to introduce redundancy back to the bitstream to strengthen it against channel errors. The question is how we can minimize the redundancy to achieve maximum protection for the bitstream. Traditional mode-selection approaches, such as the random intra-updating (RIU) technique as discussed in Section 2.4.4, achieve it by selectively intra-refreshing some of the MBs or frames. However, an intra-coded MB or frame requires much larger bit resource than an inter-coded MB or frame. We need to reconsider if there are other economic ways that can give the same or better result.

We have suggested in the above section that the MBRM approach can be an alternative. To minimize the overhead, the MBRM as suggested in Fig. 6.2 is the shortest and unique one within the MB layer. It includes only the MB number indicator and relies on the PSC/GBSC for providing supplementary information to decode an MB. Hence, the resource required for the proposed MBRM is greatly reduced as compared with the other RMs such as PSC/GBSC and the one suggested in [157]. Besides, for many video monitoring and conferencing systems having fixed camera setting, a large portion of the frame content indeed is static background. If it is coded as P-frames, the static background will often be coded in skipped mode with the current H.263 standard. Since the proposed

MBRM is only attached to the coded MBs, the overhead required for the introduction of MBRM is limited for these applications. Besides, the number of MBRMs required can be further reduced by coding the MBs with irrelevant content in skipped mode as suggested in [159].

To investigate the overhead introduced by MBRM as compared with the traditional mode-selection approach, such as RIU, a simulation for both approaches is carried out using two standard sequences, Claire and Car-phone. For both sequences, 30 frames are encoded at 10 frames/sec and all frames are inter-coded excepted that the first frame is intra-coded. The bitstreams are then fed to an error-prone channel with BER equals to 10^{-4}. For both approaches, we conceal the damaged MBs by directly copying the decoded MBs at the same spatial positions in the previous frame. The decoded results of both approaches are compared. Table 6.1 shows the overhead requirements and Fig. 6.4 illustrates the resulted PSNRs (Y-component only) of the reconstructed frames achieved by the two approaches.

Encoding rate / Method	Claire		Car-phone	
	Bit rate (Kbits/sec)	Overhead (Kbits/sec)	Bit rate (Kbits/sec)	Overhead (Kbits/sec)
Original H.263 (with GBSC)	42.11	0	71.10	0
RIU (with GBSC) (Remark: An MB will be intra-refreshed if it is not done so in the previous 10 frames)	54.14	12.03	86.71	15.61
MBRM (GBSC + MBRM)	48.23	6.12	83.23	12.13

Table 6.1: Comparison of overhead requirements of the RIU approach and the MBRM approach.

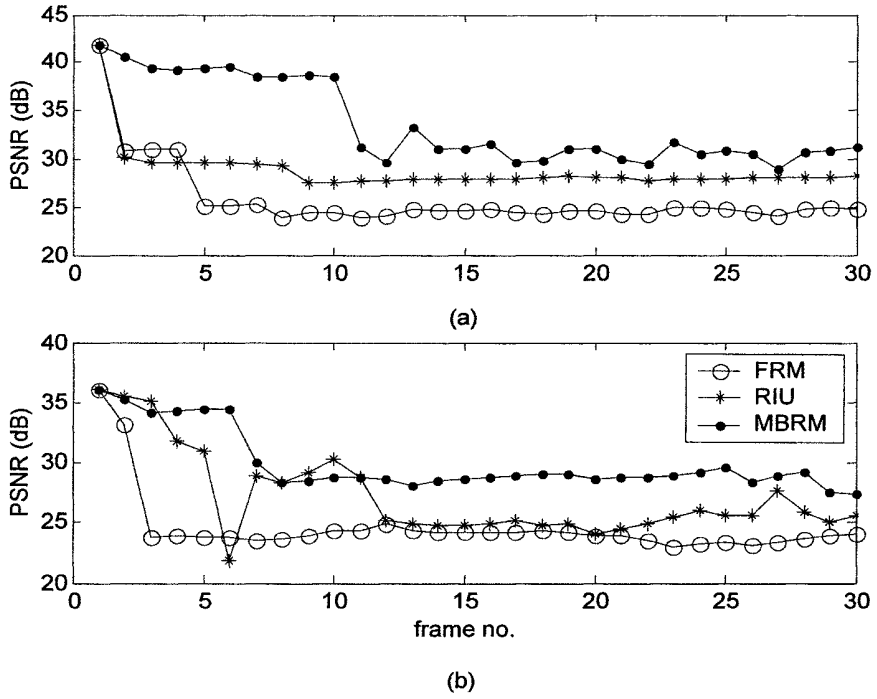


Figure 6.4: Resulted PSNRs against the decoded frames for the sequences: (a) Claire and (b) Car-phone at BER = 10^{-4}.

The results show that the overhead required for MBRM is much less than RIU while the MBRM approach achieves higher PSNR for both sequences. Thus, as shown in this simulation, the overhead introduced by MBRM is indeed not significant comparing with the commonly used RIU approach. However, for higher motion videos, e.g. Car-phone, intra-refreshing sometimes can be more useful than just concealment since it really corrects the error. That is why the RIU approach gives better PSNR in a few decoded frames than the MBRM approach for Car-phone.

The above results show that the selection of MBRM or RIU is dependent on the video content. RIU will be too expensive if the video motion is not that vivid. MBRM is a good candidate for low to medium motion videos. However, the above conclusion is made assuming that rate is unlimited. If rate is the only concern than the picture quality, both approaches should not be used indeed. Hence, for any video, there should be an optimum

combination of error resilient operating modes (MBRM, RIU, or simply none) that allows the best picture quality at the required rate constraint. To this problem, a RD-based mode-selection algorithm is proposed in the next section.

6.3.2 Optimal GOB-based mode-selection algorithm

We have shown in the last section that the proposed MBRM gives the same protection to the bitstream as the mode-selection scheme but with much lower overhead. However, using MBRM alone can only regain synchronization and relieve the effect of error by concealment. The error cannot be corrected if not intra-refreshed. In this section, a new mode-selection algorithm is proposed to allow an optimum use of MBRM and coding mode selection scheme.

First let us recall that in H.263, an MB can be encoded in either *skip*, *inter*, or *intra* mode. If MBRM is further considered, there are five modes to encode an MB: *skip*, *inter*, *inter with MBRB*, *intra*, and *intra with MBRM*. To independently select the best coding mode of an MB can be rather difficult due to the inter-dependence between the quantization value Q of adjacent MBs in a GOB. In H.263, an MB can be encoded using a differential quantization value ranged from –2 to 2 with reference to its previous MB [12]. When we consider the coding mode of an MB, we must also consider the effect to its following MBs. If there is totally L MBs in a GOB, it requires to search over 5^L possibilities of coding modes for all possible Q within the range {1, 2, ..., 31}. It needs an extremely high computational complexity. To reduce the complexity, we suggest to let all MBs within a GOB to have the same Q (as suggested in [160]) and perform mode-selection among GOBs rather than MBs. By doing so, the computational complexity can be much reduced as no dependency problem exists between GOBs. Besides, as the dependency between adjacent MBs is no longer exist (since all MBs have the same Q), an MBs can be reconstructed accurately without considering whether its former MBs are correctly received or not.

As far as GOB mode-selection is concerned, three coding modes covering different levels of protection are suggested: *normal* mode, *local inter-resynchronization* mode, and *local intra-resynchronization* mode. The *normal* mode encodes a GOB following the same strategy as the traditional H.263 without any insertion of MBRM. The *local inter-resynchronization* mode encodes the MBs in a GOB in the same way as the *normal* mode but attaches MBRM to each coded MBs. To achieve intra-refreshing for the decoded frames, RIU is integrated into the *local inter-resynchronization* mode and results in the *local intra-resynchronization* mode. The RIU approach intra-refreshes an MB if it has not been refreshed in the previous n frames [12]. The *local intra-resynchronization* mode allows to intra-code some of the MBs within a GOB according to RIU and attaches MBRM to each coded MB. For this mode, all MBs in a frame are divided into several groups randomly. Intra-refreshing is performed among one group of MBs in a frame according to RIU. In such case, each GOB contains MBs that belong to different groups. The *local intra-resynchronization* mode will be considered only if any MB in a GOB is supposed to be intra-coded by RIU. If the *local intra-resynchronization* mode is chosen, the MBs that belong to a particular group in that GOB are allowed to be intra-coded. If this mode considered but not chosen to encode a GOB, the refreshment period for those MBs supposed to be intra-coded will be reset. By adopting the suggested coding modes, the number of possible structures that we need to search for each Q for each GOB is reduced from 5^L to 3 only.

To determine the best coding mode of a GOB, we use the reconstruction distortion of that GOB as a criterion. To estimate the distortion, we need to consider the effects of channel BER P_e and the method of error concealment adopted. In our case, we conceal the corrupted MB as discussed in Section 6.3.1.*i*. Let M_{ij} denotes the j-GOB containing a set of MBs in frame-i and the expected total distortion for that GOB encoded using the *normal* mode is denoted as $D(M_{ij},normal)$. In *normal* mode, no MBRM is introduced hence the

whole GOB will be concealed if any of the MBs is found to have error. Exact evaluation of $D(M_{ij},normal)$ requires to consider all possible error conditions when reconstructing the GOB in previous frames which leads to an extremely high computation complexity. Fortunately, we have shown in Section 5.4 that the average distortion of a sequence of video frames can be approximated by the distortion of the I-frames involved. Hence, $D(M_{ij},normal)$ can be approximated as:

$$D(M_{ij},normal) \le (1-P_e)^{\sum_{y\in M_{ij}} K_y} \left(\sum_{y\in M_{ij}} E\left\{ \left| x_i^y - \hat{x}_i^y \right|^2 \right\} \right) + \left(1-(1-P_e)^{\sum_{y\in M_{ij}} K_y} \right) \left(\sum_{y\in M_{ij}} E\left\{ \left| x_i^y - \hat{x}_{I_y}^y \right|^2 \right\} \right) \quad (6.1)$$

where

$$K_y = \sum_{(k,p)\in F(y)} b_k^p .$$

In Eq. (6.1), $E\{x\}$ refers to the mean of x; x_i^y is the MB (a matrices of pixels) locates at spatial location y and $\hat{x}_i^y$ is the error-free reconstruction of x_i^y in frame-I; I_y is the frame number of which MB-y is last intra-coded and $\hat{x}_{I_y}^y$ is the reconstruction of the last intra-coded MB at position y, and b_k^p is the total number of bits to encode MB-p of frame-k. Due to motion compensation, an MB is predicted from one or more MBs in previous frames. $F(y)$ refers to a function that gives the number of all MB(s) of the previous frame(s), starting from I_y, that will affect $\hat{x}_i^y$. In practice, the function F can be implemented by keeping a table for each frame that records the size of each MB in that frame and its dependence with respect to the MBs in the previous frame(s).

For both the *local inter/intra-resynchronization* modes, resynchronization can possibly be performed for each MB as MBRM is attached. Similar to Eq. (6.1), $D(M_{ij}, local_{Inter})$ can be approximated by the following model:

$$D\left(M_{ij}, local_{Inter}\right) \le \sum_{y \in M_{ij}} \left(\begin{array}{l} \left(1-P_e\right)^{K_y} E\left\{\left|x_i^y - \hat{x}_i^y\right|^2\right\} \\ \qquad + \left(1-\left(1-P_e\right)^{K_y}\right) E\left\{\left|x_i^y - \hat{x}_{I_y}^y\right|^2\right\} \end{array} \right). \tag{6.2}$$

On the other hand, $D(M_{ij}, local_{Intra})$ can be approximated by the following model:

$$\begin{aligned} D\left(M_{ij}, local_{Intra}\right) \le & \sum_{k \in M_{ij}} \left(\left(1-P_e\right)^{b_i^k} E\left\{\left|x_i^k - \hat{x}_i^k\right|^2\right\} + \left(1-\left(1-P_e\right)^{b_i^k}\right) E\left\{\left|x_i^k - \hat{x}_{I_y}^k\right|^2\right\} \right) \\ & + \sum_{l \in M_{ij}} \left(\left(1-P_e\right)^{K_l} E\left\{\left|x_i^l - \hat{x}_i^l\right|^2\right\} + \left(1-\left(1-P_e\right)^{K_l}\right) E\left\{\left|x_i^l - \hat{x}_{I_y}^l\right|^2\right\} \right) \end{aligned} \tag{6.3}$$

where $k \in$ intra-coded MBs and $l \in$ inter-coded MBs in M_{ij}. The problem now is to select between the three coding modes to minimize the total distortion for frame-i, D_i, subject to a given rate constraint, R_c. That is:

$$\min_{\{\theta\}} \min_{\{Q_\theta\}} \sum_{j \in frame-i} D\left(M_{ij}, \theta, Q_\theta\right) \text{ Subject to } \sum_{j \in frame-i} R\left(M_{ij}, \theta, Q_\theta\right) \le R_c \tag{6.4}$$

where $R(M_{ij}, \theta, Q_\theta)$ is the coding rate associates with $D(M_{ij}, \theta, Q_\theta)$; $D(M_{ij}, \theta, Q_\theta)$ is the total estimated distortion of the j-GOB of frame-i using the coding mode combination set θ and quantization step Q_θ with:

$\theta = \{\theta_1, \theta_2, \ldots, \theta_n\}$ and $Q_\theta = \{Q_1, Q_2, \ldots, Q_n\}$

for n = number of GOB in the frame

where θ_k and Q_k are the coding mode and the Q for the k-GOB of that frame. This constrained problem can be rewritten as an unconstrained Lagrangian formulation as:

$$J\left(i, \theta, Q_\theta\right) = \sum_{all\ j \in frame-i} \left[D\left(M_j, \theta, Q_\theta\right) + \lambda R\left(M_j, \theta, Q_\theta\right)\right] \tag{6.5}$$

where λ is Lagrange multiplier. By finding the best θ and Q_θ for a λ that satisfies the given rate constraint R_c, the cost function in Eq. (6.5) can be minimized as:

$$\min_{\theta}\min_{Q_\theta} J(i,\theta,Q_\theta).$$

As the distortion and the rate of a given GOB have no dependency with other GOBs in the same frame, Eq. (6.5) can be simplified and rewritten as:

$$\min_{\theta}\min_{Q_\theta} J(i,\theta,Q_\theta) = \sum_{j=1}^{n} \min_{\theta_j}\min_{Q_j} J(M_{ij},\theta_j,Q_j) \tag{6.6}$$

where

$$J(M_{ij},\theta_j,Q_j) = D(M_{ij},\theta_j,Q_j) + \lambda R(M_{ij},\theta_j,Q_j).$$

From Eq. (6.6), the optimization problem in Eq. (6.5) can be minimized by independently selecting the best coding mode and Q for each GOB in the frame for a specific λ that satisfies R_c. The optimal values of λ, θ and Q_θ that minimize Eq. (6.6) can be obtained by using convex searching algorithm as in [100, 161]. To further reduce the computation complexity, some previous proposed RD models [146, 151] can be applied. The proposed fast mode-selection algorithm can be summarized as the following steps:

1. Generate target bit rate, R_c, for each encoded frame using the TMN-8 [11] rate control scheme. (Other frame-based rate control schemes, for example, those suggested in [147, 148, 149, 150], can also be applied.)
2. Perform mode-selection based on R_c using Eq. (6.6) for the coding modes as suggested in Eqs. (6.1)-(6.3).
3. Encode the current frame using the resulted set of coding modes and quantization steps.

6.3.3 Application of the proposed scheme in real systems

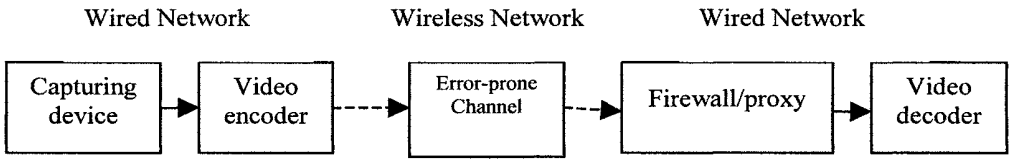


Figure 6.5: Typical architecture for one-way video transmission system over combined wireless and wired channels.

In this section, we show that the proposed mode-selection algorithm can be easily integrated into practical applications. Fig. 6.5 shows the typical architecture of a one-way video communication system, from a remote site to a local site over a combined wireless and wired channel, for many video conferencing or remote monitoring applications.

As shown in Fig. 6.5, a firewall/proxy is introduced between the decoder and the wireless network for security or stream sharing/caching purposes. For error resilience, we suggest to integrate also a transcoder into the firewall/proxy. The operation of the system is straightforward. On the encoder side, the proposed mode-selection algorithm is carried out to generate a robust bitstream according to the channel BER.

Instead of sending the bitstream directly to the decoder, a transcoder in the firewall/proxy is used to assay the bitstream and revise the damaged part if necessary. The transcoder performs the following checks to detect possible errors:

1. Invalid codeword for all VLC tables.
2. Invalid position for the RM.
3. Invalid coded MB number within the GOB.
4. The DCT coefficients in a block that exceeds 64.

The transcoder searches for the GBSC/PSC as well as the MBRM and records the corresponding MB number when it performs the above checks. If error is detected, it starts

the resynchronization and the concealment as discussed in Section 6.3.1.*i*. After removing the attached MBRMs, the resulted standard-conformed bitstream is then sent to the decoder through the wired channel. It should be noted that the proposed transcoder need not decode the bitstream entirely but only assays it up to the syntax level. Thus, it requires much lower computational effort, for instance, compared with that in [162].

6.4 Simulation Settings and Results

6.4.1 Simulation settings

Systematic study is carried out to evaluate the performance of the proposed mode-selection algorithm based on the transcoding system discussed above. Four QCIF sequences: Claire, Car-phone, Salesman, and another sequence called Traffic, each of them composes of 100 frames, are encoded at 10 frames/sec in the evaluation. The sequence Traffic is captured from a real road traffic monitoring system and its content is shown in Fig. 6.6. The four sequences are selected due to their relevance to the kinds of videos that are commonly found in video conferencing or monitoring systems. We compare the proposed approach to the traditional FRM and RIU approaches with an error-prone channel at typical BER ranged from 10^{-3} to 10^{-5}. For RIU, we arbitrarily set the refreshment period to 10. Besides, FRM is used for each GOB. We apply the same error concealment method (as mentioned in Section 6.3.1.*i*.) to the mentioned approaches such that they are competed under the same basis.

For a similar compression rate, the average PSNR (Y-component only) of the reconstructed frames, with respect to the original sequences, is computed from 10 independent runs for all approaches. The channel errors are emulated by randomly modifying some data bits of the bitstream using a random noise generator. All frames are inter-coded except that the first frame is intra-coded for each sequence. We assume that the

first frame of each sequence is received correctly. To clearly see the optimality of the proposed mode-selection algorithm, we further simulate two approaches for comparison: 1) FRM with MBRM and 2) RIU with MBRM. For both approaches, MBRMs are added to each coded MB without optimization. In all optimizations, bisection search is applied to estimate the optimal parameter set in Eq. (6.6).

Figure 6.6: Testing sequence Traffic.

6.4.2 Simulation results and discussions

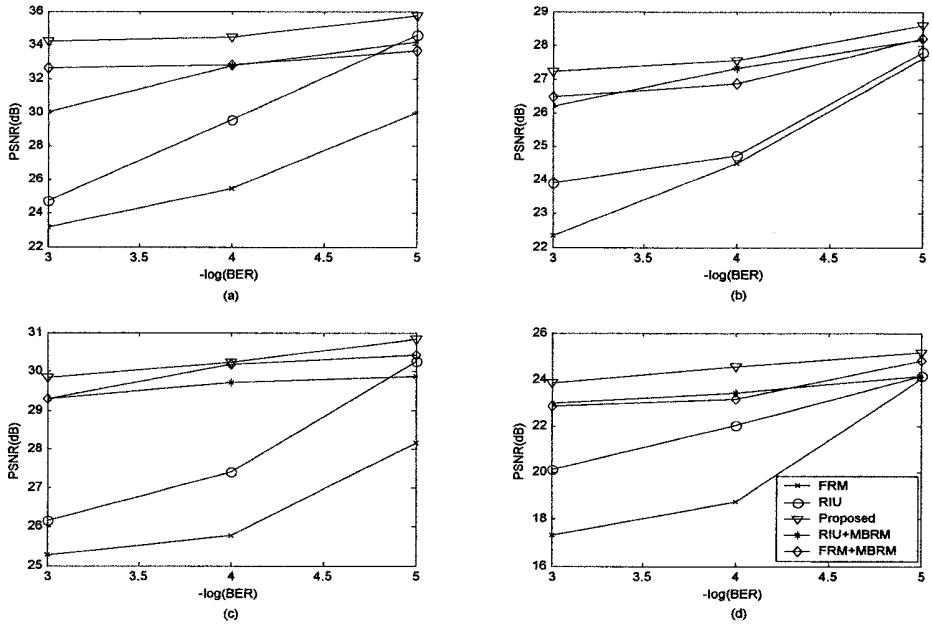


Figure 6.7: Comparison of the resulted PSNR between the proposed mode-switching approach, the well- known FRM with and without MBRM and RIU with and without MBRM for the test sequences: (a) Claire encoded at 40K bits/sec; (b) Car-phone encoded at 48K bits/sec; (c) Salesman encoded at 40K bits/sec, and (d) Traffic encoded at 108K bits/sec, subject to the channel at BER ranged from 10^{-5} to 10^{-3}.

Fig. 6.7 shows the performance of the mentioned approaches subject to different levels of BER for all sequences encoded at similar bit rate. It is shown in Fig. 6.7 that, for channel BER ranged from 10^{-5} to 10^{-3}, the resulted PSNRs of the approaches having MBRM significantly outperform those without MBRM. Particularly, when the channel BER increases from 10^{-5} to 10^{-3}, the PSNR loss for the approaches having MBRM can usually be kept within 2 dB while the other approaches give 5-8 dB loss. These unbiased results give solid support to the proposed MB level error detection and resynchronization. When comparing the approaches using MBRM but without optimization, the proposed optimal mode-selection approach usually gives better resulted PSNRs. It is noted from Fig. 6.7 that,

in some cases, good results can be obtained by solely using the FRM with MBRM without any intra-refreshing. This happens particularly when transmitting the low-motion sequences, like Claire, through low BER channel as good quality decoded frames can be obtained by error concealment.

Figs. 6.8 and 6.9 show the resulted PSNRs of the reconstructed frames for Claire and Traffic with BER ranged from 10^{-5} to 10^{-3}, respectively. From Figs. 6.8 and 6.9, it can be seen that the proposed approach provides steady performance over wide range of BER comparing with the other mentioned approaches across frames. Particularly, when the BER increases, the difference in the resulted PSNRs for the approaches with and without MBRM becomes more significant. It is because more corrupted data have to be concealed at high BER and this leads to a serious information loss to the decoded frames. Intra-refreshing some MBs can recover some damaged parts and hence the RIU approach provides better PSNRs than the FRM approach in some cases. The proposed approach shortens the resynchronization region such that rigorous error detection is provided and hence, the amount of data needs to be concealed is minimized. As a result, the proposed approach gives steady video quality over wide range of BER for both the standard sequences and real system captured sequence.

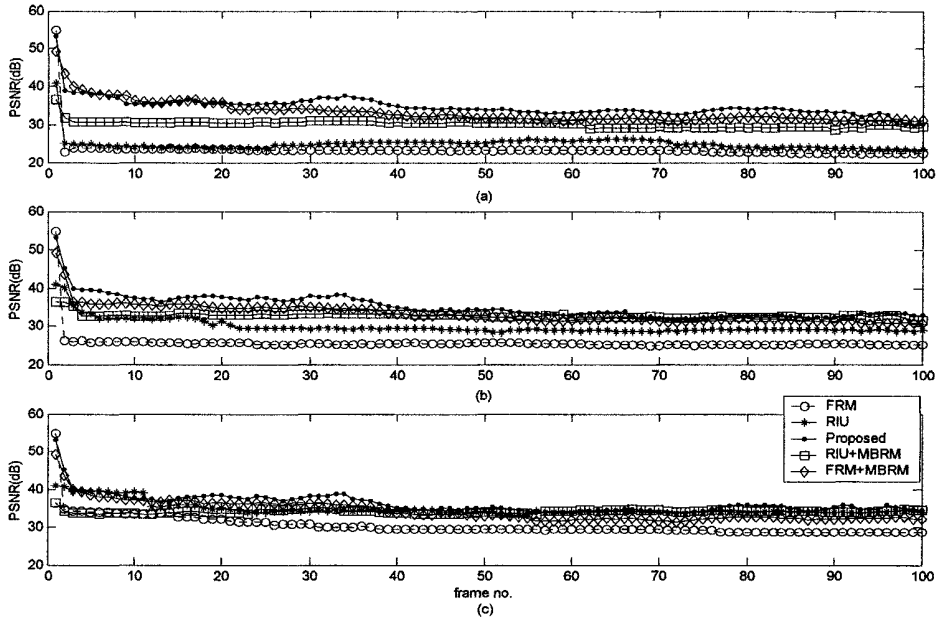


Figure 6.8: Resulted PSNRs against the decoded frames for Claire encoded by the mentioned approaches at BER: (a) 10^{-3}; (b) 10^{-4} and (c) 10^{-5}.

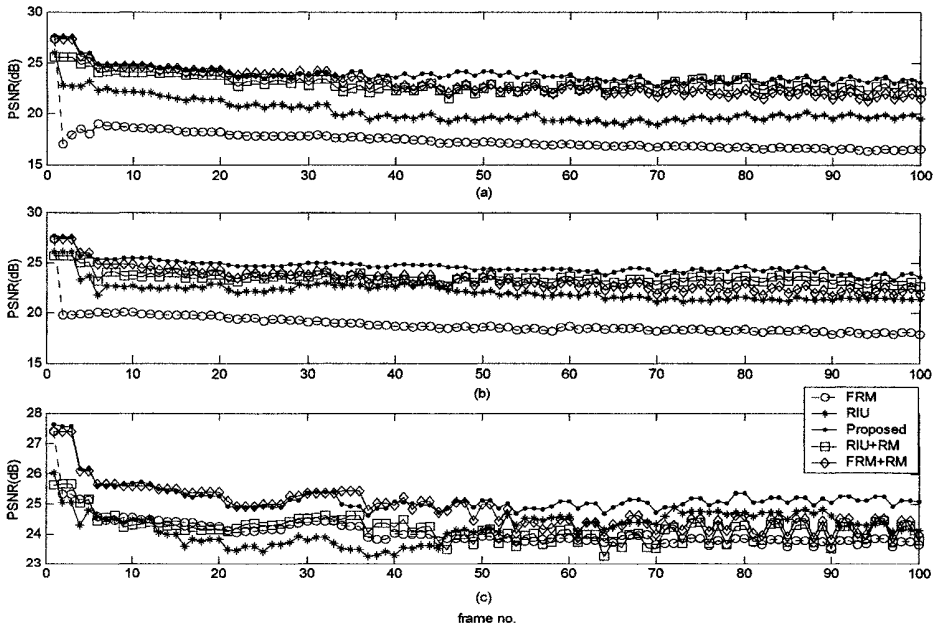


Figure 6.9: Resulted PSNRs against the decoded frames for Traffic encoded by the mentioned approaches at BER: (a) 10^{-3}; (b) 10^{-4} and (c) 10^{-5}.

6.5 Summary

In this chapter, we have given a brief review on the uses of mode-selection and RM for error resilient video coding. To improve the performance of error isolation and resynchronization in case of error, a novel RM called MBRM is proposed. Furthermore, in order to maximize the protection and the coding efficiency of the transmitted bitstream, a new GOB-based mode-selection approach is proposed by jointly optimizing the use of RM in bitstream generation. The proposed optimal GOB-based mode-selection algorithm generates robust bitstreams with strong protection from channel errors down to the MB level. Simulation results show that the proposed approach isolates the errors effectively such that the resynchronization process can be done instantaneously. The high average PSNR exhibited in the decoded frames fully justifies the advantages of the proposed approach.

Chapter 7 Conclusions

7.1 General Conclusions

In this work, we have investigated different approaches for image enhancement and robust video communication. In fact, techniques for image enhancement and reliable source coding scheme are two critical factors to achieve high quality image and video communications. The theme of this work mainly concerns the aspects related to these two factors. We carried out research works in three different areas, namely, efficient blind image restoration, image noise reduction, and error resilient video coding. In this section, we draw a general conclusion to this work.

Firstly, we have detailed the design of an efficient algorithm for blind image restoration using DPRT in Chapter 3. We show that, by using DPRT, the original 2-D GCV approach can be converted to some 1-D ones when it is applied to the problem of blind image restoration. This conversion reduces not only the memory requirement but also the computation time to a great extent in actual implementation. The proposed DPRT-based approach consistently uses less arithmetic operations than the 2-D GCV approach while the quality of the restored image is not inferior in both objective and subjective measures. The overall saving can be as much as 96%. This improvement is mainly due to the replacement of the 2-D FFT operations by some 1-D FFTs. The proposed approach is only a fast algorithm to the original 2-D GCV algorithm. Thus, it shares the same advantages and limitations in applications.

To further cope with the degradation introduced by the image acquisition process, the problem of coding noisy acquired images has been investigated in Chapter 4. In particular, we show that the quantization process can partially remove the noise component in a noisy image. We also explain why there is an optimal operating point when using the MPEG-4 still texture image coder to compress a noisy image. We show that the visual quality at the optimal point is greatly degraded by the so-called "cross" shape artifact due to the impulse noise induced by the quantization of the residual noise. An adaptive thresholding technique is then proposed to remove the residual noise by considering the pdf of the wavelet coefficients. The proposed approach selects threshold values by carefully considering the quantization problem of the encoder which is often ignored in other generic denoising approaches. By using the proposed technique, the "cross" shape noise in the reconstructed image is greatly reduced. The resulted PSNRs of the decoded images are also improved by 1 to 2 dB as compared with the traditional MPEG-4 still texture image coder with similar compression ratio. The proposed technique is straightforward and easy to implement. It requires only a slight modification to the coder while the decoder remains unchanged and is particularly suitable for the real time low bit rate image processing systems under noisy environment.

We have also studied the problem of source coding for robust video transmission. To this problem, we first study the RD relationship for typical block-based compressed video. We then investigate efficient error resilient coding scheme for robust video communications.

A comprehensive analysis on the RD relationship for block-based video codec has been provided in Chapter 5. We show that the average distortion of a GOP is dominated by the quantization error among the I-frames involved in the GOP. We also show that the distortion-quantization relationship, $D(Q)$, can be approximated by a linear model in terms of the quantization step size, Q, using regression analysis. Based on the proposed $D(Q)$

model, we derive the rate-quantization model, $R(Q)$, accordingly. The parameters of the proposed models are estimated using a least-square optimization approach. The estimated data are compared with the actual measurements and promising results are obtained. The proposed GOP-based RD model is different from the generic frame-based RD models that it determines the RD relationship for a GOP rather than an individual frame. Hence, the proposed approach need not carry out RD estimation for each frame when determining the RD relationship for a video sequence and hence requires less computational complexity.

Finally, in Chapter 6, we have discussed the error resilient coding scheme for robust video communications. To improve the performance of error isolation and resynchronization in decoders in case of error, a novel RM called MBRM is proposed in this chapter. Furthermore, in order to maximize the protection and the coding efficiency of the transmitted bitstream, a new GOB-based mode-selection approach is proposed by jointly optimizing the use of MBRM in bitstream generation. The proposed mode-selection algorithm estimates the overall distortion for each GOB in a decoded frame by carefully considering the effects of quantization, error propagation, and error concealment. It simplifies the prediction of the overall distortion for each GOB by adopting the GOP-based $D(Q)$ model proposed in Chapter 5 and then performs optimal mode-selection for each GOB in a video frame. The algorithm generates robust bitstream with strong protection from channel errors down to the MB level and isolates errors effectively such that the resynchronization process can be done instantaneously.

On the whole, we have developed four major approaches to improve the robustness of image and video communications. The proposed efficient DPRT-based blind blurred image restoration algorithm and wavelet-based adaptive thresholding technique offer effective ways for accurate image acquisition, while the GOP-based RD modeling technique and optimal GOB-based mode-selection algorithm can provide clear insight and source coding solution for reliable video transmission.

7.2 Future Work

The results obtained in this work are indeed small contributions to the development of robust image and video communications. It is anticipated that more work should be done, based on the theory and the results reported in this dissertation, to further improve the problem concerned. In this section, we discuss some of the possible extensions.

As discussed in Chapter 3, the proposed efficient DPRT-based approach assumes that the support of PSF is known in the restoration process. This assumption may not be valid in some practical situations. It means that a priori process may be required to estimate the support of the PSF. Some results from the work on ARMA model order estimation [124] can be used to deal with this problem. Besides, since the proposed approach is just a fast algorithm to the original 2-D GCV algorithm, it shares the same limitations in applications. For example, it can only estimate the magnitude of PSF but not the phase information during the restoration procedure. It is of particular interest to carry out extension work to tackle these two problems.

For the wavelet-based adaptive thresholding technique proposed in Chapter 4, the threshold values are selected according to a one-to-one mapping scheme between the cumulative probabilities for a given value under the pdfs of the wavelet coefficients and the zero mean Gaussian noise. In practice, it can be implemented by using the technique of table look-up. However, the memory requirement for keeping the tables highly depends on the precision of the threshold values. To enhance the performance of the proposed approach, it is suggested to formulate the two mentioned cumulative probabilities and to express the relationship between the required threshold values and the corresponding quantization step sizes using a single formula. By this means, the required threshold value can be calculated directly for a given quantization step size and noise variance without referring to any table.

In Chapter 5, a GOP-based RD modeling approach has been proposed. The proposed $D(Q)$ and $R(Q)$ models are derived to describe the average RD relationship for a GOP or even the whole compressed sequence that involves a finite number of GOP segments. Preliminary results show that there is a strong correlation between the RD behaviors of the compressed sequences encoded with different GOP structures. It is desirable to investigate this correlation such that the RD behaviors of different GOP structures can be predicted from each other. Besides, it is further suggested to extend the proposed models to the GOP that involves the bi-directional encoded frame (B-frame) even we believe that B-frame would not make great impact on the $D(Q)$ function of a GOP.

As for the GOB-based mode-selection algorithm, the optimal coding mode and quantization step size of each GOB in a video frame are obtained by repeating the encoding and decoding procedures several times. An extension work is recommended to apply RD modeling technique to simplify the optimization process of the proposed approach. Besides, it is also desirable to record the computational time of the proposed mode-selection algorithm, or to analyze further its complexity.

In fact, to determine the RD behavior between the encoder and the decoder in error resilient image or video compression and transmission is a very complex problem. In order to ensure the robustness of a RD model, we need to carefully consider various factors affecting the RD behavior. For example, as discussed in Chapter 6, the selection of coding mode and the placement of RM impinge on the RD behavior directly, especially in error resilient coding. Besides, it is also an interesting topic to integrate the image enhancement techniques into the compression process such that they can be done in a single step. For example, instead of our proposed adaptive thresholding technique, one can also consider to perform image restoration in the compression process. There is still ample scope for much challenging work in these areas.

Bibliography

[1] G. A. Baxes. *Digital Image Processing: Principles and Applications*. John Wiley & Sons, Inc., New York, 1994.

[2] V. K. Madisetti and D. B. Williams. *The Digital Signal Processing Handbook*. CRC Press, Boca Raton, FL., 1998.

[3] H. I. Cheng. *Image Compression: A 2-Dimensional Run-Length Coding*. Dissertation Thesis (Ph.D.)-University of Missouri-Columbia, 1983, University Microfilms International, Ann Arbor, Mich., 1986.

[4] D. Salomon. *A guide to data compression methods*. Springer, New York, 2002.

[5] A. Drozdek. *Elements of Data Compression*. Brooks/Cole Thomson Learning, Pacific Grove, Calif., 2002.

[6] K. Sayood. *Introduction to Data Compression*. Second Edition. Morgan Kaufmann Publishers, San Diego, Calif., 2000.

[7] G. K. Wallace, "The JPEG still picture compression standard," *Commun. ACM,* vol. 34, pp.30-44, April 1991.

[8] ISO/IEC JTC1/SC29/WG01 Adhoc Group on JPEG-2000 Verification Model, "JPEG-2000 verification model version 5.2," *WG1 N1422*, August 1999.

[9] J. M. Shapiro, "Embedded image coding using zero-tree of wavelet coefficient," *IEEE Trans. on Signal Processing*, vol. 41, pp. 3445-3462, December 1993.

[10] J. M. Shapiro, "Image coding using the embedded zerotree wavelet algorithm," *Mathematical Imaging, SPIE*, vol. 2034, pp.180-193, 1993.

[11] ITU-T, "Video coding for low bitrate communication," *ITU-T Recommendation H.263, Version 1*, 1996.

[12] ITU-T, "Video coding for low bitrate communications," *ITU-T Recommendation H.263. Version 2 (H.263+)*, 1998.

[13] ISO/IEC JTC1/SC29/WG11, "Information technology – Generic coding of audio-visual objects. Part 2: Visual," *Draft ISO/IEC 14496-2, Version 1, ISO/IEC*, 1998.

[14] D. Deloddere, W. Verbiest and H. Verhille, "Interactive video-on-demand," *IEEE Communication Magazine*, pp. 68-80, May 1994.

[15] R. Steinmetz and K. Nahrstedt. *Mulitmedia Computing, Communications & Applications*. Prentice Hall, 1996.

[16] Philips and Sony Corporation. *System Description Compact Disc Read Only Memory*, Yellow Book, 1985.

[17] R. E. Blahut. *Theory and Practice of Error Control Codes*. Addison-Wesley, 1983.

[18] W. W. Peterson. *Error Correcting Codes*. MIT Press, 1961.

[19] J. Anderson and S. Mohan. *Source and Channel Coding: An Algorithmic Approach*. Kluwer Academic Publishers, 1991.

[20] S. G. Wilson. *Digital Modulation and Coding*. Prentice Hall, Upper Saddle River, N.J., 1996.

[21] M. Bertero and R. Boccacci. *Introduction to Inverse Problems in Imaging*. Institute of Physics Publishing, Philadelphia, Pa., 1998.

[22] W. Kropatsch, R. Klette, and F. Solina. *Theoretical Foundations of Computer Vision*. Wien and Springer, New York, 1996.

[23] N. Saitoh, K. Kuroki, and K. Kurosawa, "Estimation of out-of-focus blur and its effectiveness for deblurring," *Proceedings, SPIE, Investigative Image Processing*, vol. 2942, pp. 2-9, Boston, Massachusetts, November 1996.

[24] R. Baldock and J. Graham. *Image Processing and Analysis: A Practical Approach*. Oxford University Press, New York, 2000.

[25] O. K. Al-Shaykh and R. M. Mersereau, "Lossy compression of noisy image," *IEEE Trans. on Image Processing*, vol. 7, no. 12, pp. 1641-1652, December 1998.

[26] Y. Wang, J. Ostermann, and Y. Q. Zhang. *Video Processing and Communications.* Prentice Hall, Upper Saddle River, N.J., 2002.

[27] A. K. Katsaggeolos. *Digital Image Restoration.* Springer-Verlag, New York, 1991.

[28] H. C. Andrews and B. R. Hunt. *Digital Image Restoration.* Prentice Hall, Englewood Cliffs, HN, 1977.

[29] B. R. Hunt, "The Application of constrained Least-Squares estimation to image restoration by digital computer," *IEEE Trans. on computer*, vol. 22, pp. 805-812, September 1973.

[30] M. R. Banham and A. K. Katsaggelos, "Digital image restoration," *IEEE Signal Processing Magazine*, vol. 14, pp. 24-41, March 1997.

[31] D. Kundur and D. Hatzinakos, "Blind image deconvolution," *IEEE Signal Processing Magazine*, vol. 13, pp. 43-64, May 1996.

[32] D. Kundur and D. Hatzinakos, "Blind image deconvolution revisited," *IEEE Signal Processing Magazine*, vol. 13, pp. 61-63, November 1996.

[33] H. Kaufman and A. M. Tekalp, "Survey of estimation techniques in image restoration," *IEEE Control Systems Magazine*, vol. 11, pp. 16-24, January 1991.

[34] A. M. Tekalp, H. Kaufman and J. W. Woods, "Identification of image and blur parameters for the restoration of noncausal blurs," *IEEE Trans. on Acoustic, Speech and Signal Processing*, vol. 34, no.4, pp.963-972, August 1986.

[35] R. L. Lagendijk, A. M. Tekalp and J. Biemond, "Maximum likelihood image and blur identification: a unifying approach," *Optical Engineering*, vol. 29, no. 5, pp. 422-435, May 1990.

[36] R. L. Lagendijk, J. Biemond and D. E. Boekee, "Hierarchical blur identification," *Proceedings, IEEE Int. Conf. Acoustics, Speech and Signal Processing* (ICASSP'90), pp. 1889-1892, 1990.

[37] S. J. Reeves and R. M. Mersereau, "Optimal estimation of the regularization parameters and stabilizing functional for regularized image restoration," *Optical Engineering*, vol. 29, pp. 446-454, May 1990.

[38] S. J. Reeves and R. M. Mersereau, "Blur identification by the method of generalized cross-validation," *IEEE Trans. on Image Processing*, vol. 1, no. 3, pp.301-311, July 1992.

[39] C. M. Cho and H. S. Don, "Blur identification and image restoration using a multilayer neural network," *Proceedings, 1991 IEEE Int. Joint Conf. Neural Networks*, vol. 3, pp. 2558-2563, 1991.

[40] C. L. Nikias and A. P. Petropulu. *Higher-Order Spectral Analysis: A Nonlinear Processing Framework.* Prentice Hall, New Jersey, 1991.

[41] R. L. Lagendijk, J. Biemond and D. E. Boekee, "Identification and restoration of noisy blurred images using the expectation-maximization algorithm," *IEEE Trans. on Acoustics, Speech and Signal Processing*, vol. 38, no. 7, July, 1990.

[42] G. Wahba, "A comparison of GCV and GML for choosing the smoothing parameter in the generalized spline smoothing problem," *Annals Statistics*, vol. 13, no. 4, pp. 1378-1402, 1985.

[43] J. Astola and P. Kuosmanen. *Fundamentals of Nonlinear Digital Filtering.* CRC Press, Boca Raton, FL, 1997.

[44] M. Nachtegael, D. V. D. Weken, D. V. D. Ville, E. Kerre, W. Philips, and I. Lemahieu, "An overview of classical and fuzzy-classical filters for noise reduction," *Proceedings, IEEE Int. Fuzzy Systems Conf.*, vol. 1, pp. 3-6, 2001.

[45] M. Sonka, V. Hlavac, and R. Boyle. *Image Processing, Analysis, and Machine Vision.* PWS Publishing, Pacific Grove, CA, 1999.

[46] R. A. Peters, II, "A new algorithm for image noise reduction using mathematical morphology," *IEEE Trans. on Image Processing*, vol. 4, no. 5, pp. 554-568, May 1995.

[47] M. Nachtegael, D. V. D. Weken, D. V. D. Ville, E. Kerre, W. Philips, and I. Lemahieu, "A comparative study of classical fuzzy filters for noise reduction," *Proceedings, IEEE Int. Fuzzy Systems Conf.*, vol. 1, pp. 11-14, 2001.

[48] D. Tschumperle and R.Deriche, "Diffusion PDEs on vector-valued images," *IEEE signal Processing Magazine*, vol. 19, pp. 16-15, September 2002.

[49] I. Daubechies. *Ten Lectures on Wavelets*. Society for Industrial and Applied Mathematics, Philadelphia, Pa., 1992.

[50] Y. T. Chan. *Wavelets Basics*. Kluwer Academic Publishers, Boston, 1995.

[51] S. Mallat and L. L. Hwang, "Singularity detection and processing with wavelets," *IEEE Trans. on Information Theory*, vol. 38, no. 2, pp. 617-643, March 1992.

[52] D. L. Donoho and I. M. Johnstone, "Unconditional based are optimal based for data compression and for statistical estimation," *Applied and Computational Harmonic Analysis*, vol. 1, pp. 100-115, 1993.

[53] D. L. Donoho and I. M. Johnstone, "Ideal spatial adaptation by wavelet shrinkage," *Biometrika*, vol. 81, no. 3, pp. 425-455, 1994.

[54] D. L. Donoho, "Denoising by soft-thresholding," *IEEE Trans. on Information Theory*, vol. 41, no. 3, pp. 613-627, May 1995.

[55] M. Budagavi, W. R. Heinzelman, J. Webb, and R. Talluri, "Combining efficiency and error robustness: wireless MPEG-4 video communication on DSP chips," *IEEE Signal Processing Magazine*, vol. 17, no 1, pp.36-53, January 2000.

[56] Y. Takishima, M. Wada and H. Murakami, "Reversible variable length codes," *IEEE Trans. on Communications*, vol. 43, no.2-4, pp.158-161, February/march/April 1995.

[57] J. Wen, and J. D. Villasenor, "Reversible variable length codes for efficient and robust image and video coding," *Proceedings, Data Compression Conf.* (DCC'98), pp. 471-480, 1998.

[58] K. Imura and Y. Machida, "Error resilient video coding schemes for real-time and low-bitrate mobile communications," *Signal Processing: Image communication*, vol. 14, pp. 519-530, May 1999.

[59] I. Moccagatta, S. Soudagar, J. Laiang, and H. Chen, "Error-resilient coding in JPEG-2000 and MPEG-4," *IEEE J. on Selected Areas in Communications*, vol. 18, no. 6, June 2000.

[60] N. T. Cheng and N. G. Kingsbury, "The EREC: An efficient error resilient technique for encoding positional information on sparse data," *IEEE Trans. on Communications*, vol. 40, pp. 140-148, January 1992.

[61] D. W. Redmull and N. G. Kingsbury, "EREC: An error resilient for coding variable-length blocks of data," *IEEE Trans. on Image Processing*, vol. 5, pp. 565-574, April 1996.

[62] V. Parthasarathy, J. W. Modestino, and K. S. Vastola, "Design of a transport coding scheme for high quality video over ATM networks," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 7, no. 2, pp. 358-376, April 1997.

[63] H. T. Zheng and K. J. R. Liu, "The subband Modulation: A joint power and rate allocation framework for subband image and video transmission," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 9 no. 5, pp. 823-838, August 1999.

[64] Z. H. He, J. F. Cai, and C. W. Chen, "Joint source channel rate-disitortion analysis for adaptive mode selection and rate control in wireless video coding," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 12, no. 6, pp. 511-523, June 2002.

[65] V. A. Vaishampayan, "Design of multiple description scalar quantizer," *IEEE Trans. on Information Theory*, vol. 39, no. 3, pp. 821-834, May 1993.

[66] M. Khansari, A. Jalali, E. Dubois, and P. Mermelstein, "Low bit rate video transmission over fading channels for wireless microcellular systems," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 6, no. 1, pp. 1-11, February 1996.

[67] Q. F. Zhu, Y. Wang, and L. Shaw, "Coding and cell loss recovery for DCT-based packet video," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 3, pp. 248-258, June 1993.

[68] M. Ghanbari, "Cell-loss concealment in ATM video codes," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 3, pp. 238-247, June 1993.

[69] H. Sun and W. Kwok, "Concealment of damaged block transform coded images using projections onto convex sets," *IEEE Trans. on Image Processing*, vol. 4, no. 4, pp.470-477, April 1995.

[70] P. Haskell and D. Messerschmitt, "Resynchronization of motion compensated video affected by ATM cell loss," *Proceedings, Int. Conf. Acoustics, Speech and Signal Processing*, (ICASSP'92), vol. 3, pp. 545-548, San Francisco, CA, March 1992.

[71] Y. Wang, S. Wenger, J. Wen, and A. K. Katsaggelos, "Error resilient video coding techniques: real-time video communications over unreliable networks," *IEEE Signal Processing Magazine*, vol. 17, no 4, pp. 61-82, July 2000.

[72] B. Girod and N. Färber, "Feedback-based error control for mobile video transmission," *Proceedings of IEEE*, vol. 87, pp. 1707-1723, October 1999.

[73] J. Konrad, "Visual communications of Tomorrow: natural, efficient, and flexible," *IEEE Communications Magazine*, vol. 39, pp. 126-133, January 2001.

[74] D. P. K. Lun, T. C. Hsung, and W. C. Siu, "On the convolution property of a new discrete radon transform and its efficient inversion algorithm," *Proceedings, IEEE Int. Sym. Circuits and Systems, ISCAS'95,* vol. 3, pp. 1892-1895, Seattle, April/May 1995.

[75] T. C. Hsung, D. P. K. Lun, and W. C. Siu, "The discrete periodic radon transform," *IEEE Trans. on Signal Processing*, vol. 44, no. 10, pp. 2651-2657, October 1996.

[76] D. E. Dudgeon and R. M. Mersereau. *Multidimensional Digital Signal Processing.* Prentice Hall, Englewood Cliffs, 1984.

[77] A. K. Jain, "Advances in mathematical models for image processing," *Proceedings of IEEE*, vol. 69, no. 5, pp. 502-528, May 1981.

[78] G. H. Golub, M. Heath, and G. Wahba, "General cross-validation as a method for choosing a good ridge parameter," *Technometrics*, vol. 21, no. 2, pp. 215-223, May 1979.

[79] J. Morlet, G. Arens, I. Fourgeau, and D. Giard, "Wave propagation and sampling theory," *Geophysics*, vol. 47, pp. 203-236, 1982.

[80] J. Morlet, "Sampling theory and wave propagation," *NATO ASI Series, Issues in Acoustic signal/Image processing and Recognition, Springer-Verlag, Berlin*, vol. 1, pp. 233-261, 1983.

[81] M. J. Shensa, "The discrete wavelet transform: wedding the à trous and Mallat algorithms," *IEEE Trans. on Signal Processing*, vol. 40, no. 10, pp. 2464-2482, 1992.

[82] S. Mallat, "Zero-crossings of a wavelet transform," *IEEE Trans. on Information Theory*, vol. 37, no. 4, pp. 1019-1033, July 1991.

[83] Y. Meyer, "Un contre-example á la conjecture de Marr et á celle de S. Mallat," *Preprint*, 1991.

[84] A. Liew and D. T. Nguyen, "Uniqueness issue of wavelet transform modulus maxima representation and a least squares reconstruction algorithm," *IEE Electronics Letters*, vol. 31, no. 20, pp. 1735-1736, September 1995.

[85] A. Liew and D. T. Nguyen, "Modulus maxima using nonexpansive projections," *IEE Electronics Letters*, no. 31, no. 13, pp. 1038-1039, June 1995.

[86] Z. Cvetković and M. Vetterli, "Discrete-time wavelet extrema representatin: Design and consistent reconstruction," *IEEE Trans. on Signal Processing*, vol. 43,no. 3, pp. 681-693, March 1995.

[87] T. C. Hsung, D. P. K. Lun, and W. C. Siu, "Denoising by singularity detection," *IEEE Trans. on Signal Processing*, vol. 47, no. 11, pp. 3139-3144, November 1999.

[88] D. L. Donoho and I. M. Jonstone, "Threshold selection for wavelet shrinkage of noisy data," *Proceedings, 16th Annual int. Conf. of IEEE Engineering in Medicine*

and Biology Society. Engineering Advances: New opportunities for Biomedical Engineers, vol. 1, pp. A24-25, 1994.

[89] R. R. Coifman and D. L. Donoho, "Translation-invariant denoising," *Wavelets and Statistics*, Springer-Verlag, pp. 125-150, 1995.

[90] M. Lang, H. Guo, J. E. Odegard, C. S. Burrus, and R. O. Wells, "Noise reduction using undecimated discrete wavelet transform," *IEEE Signal Processing Letters*, vol. 3, no. 1, pp. 10-12, 1996.

[91] Y. Hui, C. W. Kok, and T. Q. Nguyen, "Wavelet shrinkage denoising using paraunitary shift-invariant filter banks," *Proceedings, IEEE Int. Sym. On Circuits and Systems* (ISCAS'97), vol. 1, pp. 185-188, 1997.

[92] M. S. Crouse, R. D. Nowak, and R. G. Baraniuk, "Wavelet-based statistical signal processing using hidden Markov models," *IEEE Trans. on Signal Processing*, vol. 46, no. 4, pp. 886-902, April 1998.

[93] G. L. Fan and X. G. Xia, "Image denoising using a local contextual hidden Markov model in the wavelet domain," *IEEE Trans. on Signal Processing*, vol. 8, no. 5, pp. 125-128, May 2000.

[94] M. E. Al-Mualla, C. N. Canagaraigah, and D. R. Bull. *Video Coding for Mobile Communications. Efficiency, Complexity, and Resilience*. Academic Press, Amsterdam, Boston, 2002.

[95] T. Sikora, "MPEG digital video-coding standards," *IEEE Signal Processing Magazine*, vol. 14, no. 5, pp. 82-100, September 1997

[96] D. Taubman, "High performance scalable image compression with EBCOT," *IEEE Trans. on Image Processing*, vol. 9, no. 7, pp. 1158-1170, July 2000.

[97] A. Skodras, C. Christopoulos, and T. Ebrahimi, "The JPEG 2000 still image compression standard," *IEEE Signal Processing Magazine*, vol. 18, no. 5, pp. 36-58, September. 2001.

[98] ISO/IEC JTC1/SC29/WG11, "MPEG-2 video test model," *MPEG93/457*, April 1993.

[99] R. Zhang, S. L. Regunathan, and K. Rose, "Video coding with optimal Inter/Intra mode switching for packet loss resilience," *IEEE J. on Sel. Areas Communications*, vol. 18 , no. 6, pp. 966-976, 2000.

[100] F. L. Léannec, F. Toutain, and C. Gillernot, "Packet loss resilient MPEG-4 complaint video coding for the Internet," *Signal Processing: Image Communication*, vol. 15, pp. 35-56, 1999.

[101] S. Regunathan, R. Zhang, and K. Rose, "Scalable video coding with robust mode selection," *Signal Processing: Image Communication*, vol. 16, pp. 725-732, 2001.

[102] S. S. Hemami and T. H. Y. Meng, "Transform coded image reconstruction exploiting interblock correlation," *IEEE Trans. on Image Processing*, vol. 4, pp. 1023-1027, July 1995.

[103] S. Aign and K. Fazel, "Temporal & spatial error concealment techniques for hierarchical MPEG-2 video codec," *Proceedings, Globecom'95*, pp. 1778-1783, 1995.

[104] J. W. Suh, and Y. S. Ho, "Error concealment based on directional interpolation," *IEEE Trans. on Consumer Electronics*, vol. 43, no. 3, pp. 295-302, August 1997.

[105] L. Atzori, F. G. B. De Natale, and C. Perra, "A spatio-temporal concealment technique using boundary matching algorithm and mesh-based warping (BMA-MBW)," *IEEE Trans. on Multimedia*, vol. 3, no. 3, pp. 326–338, September 2001.

[106] P. Salama, N. Shroff, and E. J. Delp, "A Bayesian approach to error concealment in encoded video streams," *Proceedings, Int. Conf. on Image Processing* (ICIP'1996), vol. 1, pp. 49-52, 1996.

[107] H. Sun, K. Challapali, and J. Zdepski, "Error concealment in digital simul-cast AD-HDTV decoder," *IEEE Trans. on Consumer Electronics*, vol. 18, no. 3, pp. 108-117, August 1992.

[108] N. P. Galatsanos and A. K. Katsaggelos, "Methods for Choosing the Regularization Parameter and Estimating the Noise Variance in Image Restoration and their Relation," *IEEE Trans. on Image Processing*, vol. 1, pp. 322-336, July 1992.

[109] A. K. Katsaggelos, "Iterative Image Restoration Algorithm," *Optical Engineering*, vol. 28, no. 7, pp. 35-748, July 1989.

[110] N. Fortier, G. Demoment, and Y. Goussard, "Implementation and practical comparison of two estimators of the smoothing parameter in linear image restoration," *Proceedings, Int. Conf. on Acoustics, Speech, and Signal Processing* (ICASSP'90), pp. 1905-1908, 1990.

[111] N. Nguyen, G. Golub, and P. Milanfar, "Blind Restoration/Superresolution with Generalized Cross-Validation Using Gauss-Type Quadrature Rules," *Proceedings, Thirty-Third Asilomar Conf. on Signals, Systems, and Computers,* vol. 2, pp. 1257-1261, 1999.

[112] S. R. Deans. *The Radon Transforms and Some of its Applications.* John Wiley and Sons, New York, 1983.

[113] G. Beylkin, "Discrete Radon Transform," *IEEE Trans. on Acoustics, Speech and Signal Processing*, vol. 35, no. 2, pp. 162-172, February 1987.

[114] B. T. Kelly and V. K. Madisetti, "The fast discrete Radon transform-I: Theory," *IEEE Trans. on Image Processing*, vol. 2, no. 3, July 1993.

[115] I. Gertner, "A new efficient algorithm to compute the two-dimensional discrete Fourier transform," *IEEE Trans. on Acoustics, Speech and Signal Processing*, vol. 36, no. 7 pp. 1036-1050, July 1988.

[116] E. D. Bolker, "The finite Radon transform," *Contemporary Mathematics*, vol. 63, pp. 27-50, 1987.

[117] F. Matúš and J. Flusser, "Image representation via a finite Radon transform," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 15, no. 10, pp. 996-1006, October 1993.

[118] D. P. K. Lun and W. C. Siu, "An efficient in-place computation for the two-dimensional discrete Fourier and Hartley transform," *Proceedings, Int. Sym. on Circuits and System* (ISCAS'92), pp.160-163, 1992.

[119] D. E. Dudgeon and R. M. Mersereau. *Multidimensional Digital Signal Processing.* Prentice-Hall, Englewood Cliffs, NJ, 1984.

[120] P. A. Devijver and J. Kittler. *Pattern Recognition: A Statistical Approach*, Prentice Hall, Englewood Cliffs, NJ, 1982.

[121] S. Geisser, "The Predictive Sample Reuse Method with Applications," *J. Amer. Statistical Assoc.*, vol. 27, no. 350, pp.320-328, 1975.

[122] R. L. Lagendijk, J. Biemond, and D. E. Boekee, "Identification and restoration of noisy blurred images using the expectation-maximization algorithm," *IEEE Trans. on Acoustics, Speech and Signal Processing*, vol. 38, no. 7, pp. 1180-1191, July 1990.

[123] S.C. Chan and K.L. Ho, "Polynomial transform fast Hartley transform," *Electronics Letters*, vol. 26, no. 22, pp. 1914-1916,October 1990.

[124] X. D. Zhang and Y. S. Zhang, "Determination of the MA order of an ARMA process using sample correlations," *IEEE Trans. on Signal Processing*, vol. 41, pp. 2657-2664, August 1993.

[125] ISO/IEC 144962-2, "Information Technology–Coding of audio-visual objects: visual," Comm. Draft, October 1997.

[126] W. K. Pratt, *Digital Image Processing*, Second Edition, Wiley-Interscience, 1991.

[127] B. Widrow, I. Kollar, and M. C. Liu, "Statistical theory of quantization," *IEEE Trans. on Instrumentation and Measurement*, vol. 452, pp. 353-361, April 1996.

[128] M. K. Brown, "On quantization of noisy signals," *IEEE Trans. on Signal Processing*, vol. 39, no. 4, pp. 836-841, April 1991.

[129] P. W. Wong, "Quantization noise, fixed-point multiplicative roundoff noise, and dithering," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, vol. 38, no. 2, pp. 286-300, February 1990.

[130] Y. Ephraim and R. M. Gray, "A unified approach for encoding clean and noisy sources by means of waveform and autoregressive model vector quantization," *IEEE Trans. on Information Theory*, vol. 34, no. 4, pp. 826-834, July 1988.

[131] T. R. Fischer, J. D. Gibson, and B. Koo, "Estimation and noisy source coding," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, vol. 38, no. 1, pp. 23-34, January 1990.

[132] T. Linder, G. Lugosi, and K. Zeger, "Empirical quantizer design in the presence of source noise or channel noise," *IEEE Trans. on Information Theory*, vol. 432, no. 1, pp. 612-623, March 1997.

[133] A. B. Watson, G. Y. Yang, J. A. Solomon, and J. Villasenor, "Visibility of wavelet quantization noise," *IEEE Trans. on Image Processing*, vol. 6, no. 8, pp. 1164-1175, August 1997.

[134] D. L. Donoho and I. M. Johnstone, "Adapting to unknown smoothness via wavelet shrinkage," *J. of the American Statistical Association, Theory and Methods*, vol. 90, no. 432, pp. 1200-1224, December 1995.

[135] E. P. Simoncelli and E. H. Adelson, "Noise removal via Bayesian wavelet coring," *Proceedings, Int. Conf. on Image Processing* (ICIP'96), vol. 1, pp. 79-382, 1996.

[136] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. Journal*, vol. 27, pp. 379-423, 1948.

[137] H. Gish and J. N. Pierce, "Asymptotically efficient quantizing," *IEEE Tras. on Information Theory*, vol. IT-14, pp. 676-683, September 1968.

[138] T. Berger. *Rate Distortion Theory*. Prentice Hall, Englewood Cliffs, NJ, 1984.

[139] A. Ortega and K. Ramchandran, "From rate-distortion theory to commercial image and video compression technology," *IEEE Signal Processing Magazine*, vol. 15, no. 6, pp. 20-22, November 1998.

[140] J. Ribas-Corbera and S. Lei, "Rate control in DCT video coding for low-delay communications," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 9, no. 1, pp. 172-185, February 1999.

[141] H. M Hang and J. J. Chen, "Source model for transform video coder and its application – Part I: Fundamental theory," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 7, no. 2, pp. 287-298, April 1997.

[142] H. M Hang and J. J. Chen, "Source model for transform video coder and its application - Part II. Variable frame rate coding," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 7, no. 2, pp. 299-311, April 1997.

[143] A. Ortega and K. Ramchandran, "Rate-distortion methods for image and video compression," *IEEE Signal Processing Magazine*, vol. 15, no. 6, pp. 23-50, November 1998.

[144] J. L. H. Webb and K. Oehler, "A simple rate-distortion model, parameter estimation, and application to real-time rate control for DCT-based coders," *Proceedings, Int. Conf. on Image Processing* (ICIP'97), vol. 2, pp. 13-16, 1997.

[145] K. H. Yang, A. Jacquin, and N. S. Jayant, "Normalized rate-distortion model for quantization in MC-DCT video coders," *IEE Electronics Letters*, vol. 33, no. 25, pp. 2099-2100, December 1997.

[146] T. Chiang and Y. Q. Zhang, "A new rate control scheme using quadratic rate distortion model," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 7, no. 1, pp. 246-250, February 1997.

[147] Z. He, Y. K. Kim, and S. K. Mitra, "ρ-domain source modeling and rate control for video coding and transmission," *Proceedings, Int. Conf. on Acoustics, Speech, and Signal Processing* (ICASSP'01), vol. 3, pp. 1773-1776, February 2001.

[148] Z. He, Y. K. Kim, and S. K. Mitra, "Low-delay rate control for DCT video coding via ρ-domain source modeling," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 11, no. 8, pp. 928-940, August 2001.

[149] Z. He and S. K. Mitra, "Novel rate-distortion analysis framework for bit rate and picture quality control in DCT visual coding," *IEE Proceedings – Vision, Image and Signal Processing*, vol. 148, no. 6, pp. 398-406, December 2001.

[150] Z. He and S. K. Mitra, "A unified rate-distortion analysis framework for transform coding," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 11, no. 12, pp. 1221-1236, December 2001.

[151] L. J. Lin and A. Ortega, "Bit-rate control using piecewise approximated rate-distortion characteristics," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 8, no. 4, pp. 446-459, August 1998.

[152] V. Ratnakar and M, Livny, "RD-OPT: An efficient algorithm for optimizing DCT quantization tables," *Proceedings, Data Compression Conf.* (DCC'95), pp. 332-341.

[153] Y. Lam and W. Goodman, "A mathematical analysis of the DCT coefficient distributions for images," *IEEE Trans. on Image Processing*, vol. 9, no. 10, pp. 1661-1666, October 2000.

[154] R. V. Hogg and E. A. Tanis. *Probability and Statistical inference,* Fourth Edition. Macmillan Publishing Company, New York, 1993.

[155] V. Viterbi and J. Omura. *Principles of Digital Communication and Coding.* McGraw-Hill Electrical Engineering Series, New York, 1979.

[156] R. Talluri, "Error-resilient video coding in the ISO MPEG-4 standard," *IEEE Communications Magazine*, vol. 36, no. 6, pp. 112-119, June 1998.

[157] K. Y. Yoo, "Adaptive resynchronization marker positioning method for error-resilient video transmission," *IEE Electronic Letters*, vol. 34, no. 22, pp. 2084-2085, 1998.

[158] S. H. Lee and J. K. Kim, "Optimization-based placement of resynchronization markers for error robust video transmission," *IEE Electronics. Letters*, vol. 37, no. 6, pp. 348-349, 2001.

[159] K. H. Ho and D. P. K. Lun, "Road traffic monitoring system using content-based scalable H.263 video coding scheme," *Proceedings, IEEE Pacific-Rim Conf. on Multimedia* (PCM'2000), pp. 257-260, Australia, 2000.

[160] T. Wiegand, M. Lightstone, D. Mukherjee, T. G. Campbell, and S. K. Mitra, "Rate-distortion optimized mode selection for very low bit rate video coding and the

emerging H.263 standard," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 6, no. 2, pp. 182-190, 1996.

[161] K. Ramchandran and M. Vetterli, "Best wavelet packet bases in a rate-distortion sense," *IEEE Trans. on Image Processing*, vol. 2, no. 2, pp. 160-175, 1993.

[162] G. D. L. Reyes, A. R. Reibman, S. F. Chang, and J. C. I. Chuang, "Error-resilient transcoding for video over wireless channels," *IEEE J. Selected Areas in Communications*, vol. 18, no. 6, pp. 1063-1074, 2000.