THE HONG KONG POLYTECHNIC UNIVERSITY

Singularity Detection for Regularity Scalable Image Coding

by

HO Yuk Fan

A Thesis submitted to The Hong Kong Polytechnic University

in accordance with the regulations for the Degree of

Master of Philosophy

Department of Electronic and Information Engineering

The Hong Kong Polytechnic University

October 2002

Abstract of thesis entitled
'Singularity Detection for Regularity Scalable Image Coding' submitted by HO Yuk Fan for the degree of Master of Philosophy at The Hong Kong Polytechnic University in October 2002

Scalability has been one of the important design criteria of modern image and video codec. The features such as resolution and SNR scalable coding provided in the EZW of MPEG-4 and the EBCOT in JPEG2000 are typical examples. However, the quality levels provided by the above two scalabilities do not correlate with visual perception, which means they do not particularly emphasize the display of some important features such as edges, boundaries, textures and surfaces of an image. So various feature-based progressive or scalable wavelet image coding algorithms were proposed. Nevertheless, their implementation cannot simultaneously satisfy some practical concerns such as coding efficiency and implementation complexity. In this thesis, a new scalable coding technique, namely, regularity scalable image coding, is proposed. We refer to a coding system that generates compressed bitstream in the order of the regularity of the image concerned. This algorithm can be fully embedded into any existing wavelet codec and avoid the problems that the other feature-based scalable coders encounter.

From fluid dynamics in physics to pattern recognition or computer vision in engineering, it is known that regularity (or singularity) of signals or images can be estimated from the interscale evolution of their wavelet transform. However, our regularity scalable coding algorithm is embedded into a wavelet codec, where the separable wavelet transform with decimation is applied. Therefore, the first objective of our work is to investigate the approach for estimating Lipschitz regularity from the separable wavelet transform. To avoid the error and ambiguities when tracing the

modulus maxima at coarser scales, we study the existing interscale evolution of the magnitude sums over the 'cone of influence'. Since it cannot be applied here directly to estimate the Lipschitz regularity, we determine the magnitude sums over the decimated 'cone of influence', which was not developed before. The PSNR and subjective quality results show that the coding efficiency is higher than applying resolution scalability alone.

Multiwavelet transform has been applied to image compression and denoising in the past few years. Multiwavelet transform is a generalization of the traditional single wavelet with higher multiplicity. It offers simultaneous orthogonality, symmetry and short supports, which are not possible with single wavelet transform. Due to the shorter support and higher vanishing moments of the multifilters, it generally outperforms the single wavelet transform. As the second part of our work, the singularity detection is further extended to multiwavelet systems. Again we investigate the relationship between the interscale evolution of the multiwavelet coefficients and Lipschitz regularity. The shorter supports of the multifilters introduce smaller size of 'cone of influence', which can be more clearly determined at lower scales. Hence improved results for singularity detection can be achieved. We also perform thresholding according to the interscale difference of the magnitude sum, which is complementary to the interscale ratio, in the denoising algorithm. The MSE and subjective quality results depict the performances from our theory.

Singularity detection by the Lipschitz regularity condition plays an important role in the above signal processing applications, because it fully represents the characteristics of the features of a signal or an image. We believe that it can be further applied to more signal processing applications.

Acknowledgement

I would like to take this opportunity to express my gratitude to various bodies in The Hong Kong Polytechnic University. First of all, I would like to thank my supervisor Dr. Daniel P. K. Lun for his guidance, encouragement, support and discussion, concerning to my study and the attitude of life. These become precious and important knowledge and experiences that enrich me. I am glad to thank the board of examiners, Dr. K. P. Chan and Dr. L. M. Po from The University of Hong Kong and City University of Hong Kong respectively for their attention.

I would also like to thank Dr. T. C. Hsung, who collaborated with me throughout my research study. I truly thank for his constant and kind supports, help, advice and interesting discussion.

I have to thank all the members and staffs of the Department of Electronic and Information Engineering and the Centre for Multimedia Signal Processing. They provided a creative and comfortable environment for me to work in. I thank for their friendship and encouragement. It is also my pleasure to appreciate the support from the Research and Postgraduate Studies Office in the past two years.

Of course, the prayers, unforgettable support and unforeseen concern from the brothers and sisters in the Peace Evangelical Church are greatly acknowledged. I thank my boy friend for his endless encouragement and deep love. Moreover, I thank my parents and sibling. Without them, this study could never be continued and completed.

Last, but not the least, I greatly thank the Lord Jesus Christ for His greatest love. He gave me confidence, peace and insights that helped me to overcome all the difficulties I encountered during the past two years. I dedicate this thesis to the glory of God.

Statement of Originality

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

1. The Lipschitz regularity condition for separable wavelet transform (Chapter 3, Section 3.3)

   Lipschitz regularity is characterized by the wavelet transform and the Lipschitz exponents can be determined from the magnitude of the wavelet transform across scales. As there are many advantages to embed the proposed scalable coding algorithm into a wavelet image coder, where the separable wavelet transform with decimation is applied, we need to characterize Lipschitz regularity from the separable wavelet transform. We prove the Lipschitz regularity condition for the separable wavelet transform (Appendix I).

2. The adaptive regularity scalable wavelet image coding algorithm (Chapter 3, Section 3.3)

   Since the existing WTMS over the COI cannot be directly applied to the separable wavelet transform with decimation, in this algorithm, the regularity of a signal is estimated by computing the WTMS over the decimated COI of the separable wavelet transform with decimation. This is based on the work in the first statement of originality. Then the wavelet coefficients are selected and classified to various regularity levels according to their interscale ratios (and differences) of their WTMS. The comparison with the existing feature-based scalable wavelet image coding algorithms showed that the decoded images are higher in PSNR and have better visual quality at the same bit rates. Significant improvements of the edges and textures components of a decoded image are obtained at lower bit rates. When we combine this scalable coding algorithm with resolution scalability, the bit rates of a decoded image can be greatly reduced at the same quantization level. This

scalable coding algorithm also possesses good noise robustness and it is image adaptive. It would be very useful in image browsing and retrieval applications.

3. The Lipschitz regularity condition for the multiwavelet transform (Chapter 4, Section 4.3)

   Lipschitz regularity is characterized by the interscale evolution of the single wavelet transform in terms of Lipschitz exponent. For multiwavelet transform, as the supports of the multifilters are shorter, the width of the 'cones of influence' becomes narrower and the overlappings of the singularities at coarser scales become less. Furthermore, the vanishing moments of the multifilters is sufficient to provide a better signal approximation so that the wavelet coefficients produced at fine scales become smaller. Therefore, we attempt to state and prove the relationship between the interscale evolution of the multiwavelet transform and the Lipschitz exponent in order to perform singularity detection in multiwavelet transform and verify the above nice properties of multiwavelet transform over the single wavelet transform. (Appendix II)

4. The signal denoising algorithm using the multiwavelet singularity detection (Chapter 4, Section 4.3)

   In this algorithm, we estimate the Lipschitz regularity of a noisy signal by computing the multiwavelet transform magnitude sum over the COI of the multiwavelet transform. We use the interscale ratio and difference of the modulus of the two magnitude sum components to reject wavelet coefficients which belong to the noise components. We experimented two selection approaches, they are thresholding the coefficients within the COI and thresholding the maxima coefficients inside the COI. We then simply reconstruct the denoised signal by simply applying the inverse multiwavelet transform. We found that signal denoising using the multiwavelet singularity detection is better than signal denoising using the single wavelet singularity detection.

Table of Contents

Author Publications

0.1 International journal papers

- Ho, Y.-F., Hsung, T.-C., Lun, D. P.-K. and Tam, P. K.-S. "Regularity Scalable Wavelet Image Coding Based on Singularity Detection". *IEEE Transactions on Image Processing*. (to be submitted)

- Ho, Y.-F., Hsung, T.-C., Lun, D. P.-K. and Tam, P. K.-S. "Multiwavelet Denoising Using Singularity Detection," *IEEE Transactions on Signal Processing*. (to be submitted)

0.2 International Conference/Symposium papers

- Ho, Y.-F., Hsung, T.-C. and Lun, D. P.-K. "Adaptive Regularity Scalable Wavelet Image Coding". *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*, volume 4, pages 3493-3496 (2002)

- D. P.-K. Lun, T.-C. Hsung and Y.-F. Ho, "Wavelet Singularity Detection for Image Processing," *Proceedings of the 45th IEEE International Midwest Symposium on Circuits and Systems*. (2002) (accepted)

- Ho, Y.-F., Hsung, T.-C., Lun, D. P.-K. and Tam, P. K.-S. "Multiwavelet Denoising Using Singularity Detection," *IEEE International Conference on Acoustics, Speech and Signal Processing 2003*. (to be submitted).

List of Figures

List of Tables

Notations

$\langle f, g \rangle$	Inner product.
$\|f\|$	Norm.
$f = O(g)$	Order of: there exists K such that $f \leq Kg$.
$A < +\infty$	A is finite.
$A >> B$	A is much larger than B.
h^*	Complex conjugate of $h \in C$.
$\lfloor x \rfloor$	Largest integer $n \leq x$.
i	$i^2 = -1$.

Sets

N	Natural numbers.
Z	Integers.
$\Re$	Real numbers.
$\Re^2$	$\Re \times \Re$.
$\Re^+$	Positive real numbers.
C	Complex numbers.

Signals

$f(t)$	Continuous time signal.
$f[n]$	Discrete time signal.
$\delta(t)$	Dirac distribution.
$\delta[n]$	Discrete dirac.

Spaces

C^p	p times continuously differentiable functions.
$L^2(\Re)$	Finite energy functions $\int\|f(t)\|^2 dt < +\infty$.
$L^2(\Re, C)$	Complex-valued finite energy functions.
$V \oplus W$	Direct sum of two vector spaces.

Operators

$f'(t)$	First derivative.
$f^{(n)}(t)$	Derivative of order n.
$f * g$	Convolution

Transforms

$\hat{f}$	Fourier transform.
$Wf(u,s)$	Wavelet transform.
$P_V f(u,\xi)$	Wigner-Ville distribution.

Probability

X	Random variable.
$\Pr(X)$	Probability.
$E\{X\}$	Expectation.
$Var(X)$	Variance.

Glossary

n-d	n-dimensional.
WTMM	Wavelet transform modulus maxima.
COI	Cone of influence.
WTMS	Wavelet transform modulus sum.
POCS	Projection onto convex set.
GHM	Geronimo-Hardin-Massopust.
CL	Chui-Lian.
FIR	Finite impulse response.
IIR	Infinite impulse response.
QoS	Quality of service.
EZW	Embedded zerotree wavelet.
MPEG	Moving picture experts group.
SPIHT	Set partitioning in hierarchical trees.
EBCOT	Embedded block coding with optimized truncation.
JPEG	Joint pictures experts group.
SUSAN	Smallest univalue segment assimilating nucleus.
HVS	Human visual system.
USC	University of Southern California.
SURE	Subspace rotation estimation.
MRA, VMRA	Multiresolution analysis, Vector multiresolution analysis.
QMF, MQMF	Quadrature mirror filter, matrix quadrature mirror filter.
SNR	Signal-to-noise ratio $\sum_{\forall i}(x_i)^2 / \sum_{\forall i}(x_i - \hat{x}_i)^2$.

PSNR	Peak signal-to-noise ratio $10\log_{10}\sum_{\forall i} x_i^2 / \sum_{\forall i}(x_i - \hat{x}_i)^2$.
MSE	Mean square error $\sum_{\forall i}(x_i - \hat{x}_i)^2 / N$.

Chapter 1

Introduction

1.1 Objectives of our research

Current feature-based scalable wavelet image coding algorithms cannot fulfill fully the concern of the correlation with visual perception, coding efficiency and implementation complexity. This motivates us to develop a simpler and more effective scalable wavelet image coding algorithm. The proposed scalable coding algorithm should achieve the following properties:

- The scalability is based on the visually important features of an image.
- It is image adaptive.
- It can combine with other scalable coding algorithms such as the resolution and SNR or accuracy scalabilities simultaneously, without rearrangement of bitstream.
- It maintains a reasonably high coding efficiency.

1.2 Significance of our works

To achieve the first objective, we need to quantify the regularity of an image and produce a scalable representation. Edge of different structures that appear in an image is often the most important feature which characterizes the regularity of an image. In computer vision or pattern recognition, multiscale edges of images are often detected by oriented wavelets with partial derivative of order one [4,39]. Visually important features such as edges are further quantified from the interscale evolution of the oriented wavelet transform. Visual smoothness or regularity of an image is related to the asymptotic decay of oriented wavelet transforms in the corresponding neighborhood in two directions, and the decay is controlled by the wavelet transform modulus [40,76]. Mathematically, it was shown in [40] that the local wavelet transform modulus maxima (WTMM) provide numerical procedures to compute Lipschitz exponents, which quantify the local visual smoothness of an image. Nevertheless, errors and ambiguities are introduced when tracing WTMM at coarser scales, so wavelet transform modulus sum (WTMS) was proposed [32] as a replacement of WTMM. This was applied as a post-processing technique to preserve the edges of noisy natural or tomographic images [32,89,6]. In order to maintain a high coding efficiency, as stated in the forth objective, an obvious approach is to embed the algorithm in the wavelet image coder. Whereas for image compression in

wavelet image coder, the separable wavelet transform with decimation is usually performed for the decomposition. For the implementation, the wavelet coefficients are computed with fast filter bank algorithm that cascades discrete convolutions with dilated filters and subsamplings [53,68,69,83,84,70,78,79,82]. Operating on these decimated coefficients is more convenient and less computative, as the process is embedded and less amount of coefficients is involved. WTMM and WTMS algorithms only operate on the oriented wavelet transform, which is undecimated or without any subsamplings. From [40,72,74,75,76,80], the Lipschitz regularity condition, which determines the Lipschitz exponents, was just derived for the oriented wavelet transform. With the assumption of using a jointly shiftable translation invariant wavelet transform [25,46,67], we reckon that separable wavelet transform also have a similar Lipschitz regularity condition, and this is shown in our work [30].

After the Lipschitz regularity of a separable wavelet transformed image is quantified, we seek for a regularity scalable representation for it. With a set of well designed wavelet filters, [46] shows that a 2-d wavelet transform can be jointly shiftable in position and orientation. Wavelet coefficients corresponding to different regularity levels can then be obtained by selecting wavelet coefficients at some interscale ratio thresholds, which are related to the Lipschitz exponents. There are errors and ambiguities of tracing WTMM at coarser scales, and the decimation may

clear some maxima points and make it may become inaccurate to estimate the Lipschitz regularity. The reconstruction process is also very complicated. So the WTMS algorithm becomes the ultimate solution. It is because it can consider a wider range of the effect from a singularity in the frequency domain, which is more localized, on the other side, in the spatial domain, especially at coarser scales with denser subsamplings. This better measurement of regularity is verified in [32]. However, for the separable wavelet transform, we cannot compute magnitude sum directly to estimate the Lipschitz exponents in turn to produce a regularity scalable representation of an image. 'Cone of influence' is the influential region of the response produced by an impulse bounded by the support of the wavelet kernels. It is proportional to the scale. Here we construct the decimated 'cone of influence' for computing the magnitude sums of the separable wavelet transform. Moreover, the effect of decimation to the magnitude sum is relatively less than the maxima.

There are many advantages of adopting the scalable coding algorithm into zerotree wavelet image coders. Apart from the high coding efficiency it could enjoy from the image coders, the regularity levels defined can be adaptive to images. Normalization of the interscale ratio of the magnitude sums is performed for obtaining an even progression of regularity across the scalability levels. As the magnitude of a coefficient represents the energy of the coefficient possesses, with partial ordering of

the magnitudes of the coefficients and the interscale ratios of their magnitudes or magnitude sums, we obtain an optimal arrangement of the encoding order so that lower regularity components of an image are encoded first. Since the regularity scalability can be exhibited with the resolution and SNR scalabilities simultaneously, no change in the original bitstream format is needed. Furthermore, the combination of bitstream that produces the sequence for transmission is simple. Image data is encoded and transmitted in the order of regularity evaluated from the interscale evolution of the wavelet coefficient magnitudes or their sums. Then it is undergone the zerotree structure coding, and the lower regularity components, which is more visually important, are encoded and transmitted first at certain resolution and accuracy, with higher regularity components are encoded progressively as more bits are allowed. Therefore, it provides good visual perception at very low bit rates or an emphasis on a particular feature at a certain regularity of an image. Typically, the encoding process can be stopped before or when the target bit rate is met. Similarly, the decoder can interrupt decoding at any point in the bit stream and reconstruct an image at a certain regularity level, resolution and accuracy.

1.3 Organization of this thesis

In chapter 2, we make a brief review on the wavelet transform. We begin with the history and development of the wavelet transform. We outline its important features and its realization, from fast dyadic transform to discrete filter bank implementation. We then come to the studies of the wavelet singularity detection and its applications in pattern recognition, computer vision and signal processing. As we further extend the singularity detection to the multiwavelet transform, we also briefly review the history and development of the multiwavelet transform. We outline its important features and its realization with the prefiltering.

In chapter 3, we present the proposed adaptive regularity scalable wavelet image coding algorithm. We first recall several existing feature-based scalable wavelet image coding algorithms, their features and their performances. Then we investigate the Lipschitz regularity condition for the separable discrete wavelet transform. We determine the decimated COI for the magnitude sums in the separable discrete wavelet transform. Then this is applied to our proposed algorithm. Next we present the procedures of implementing the proposed scalable coding algorithm, the results and the comparison with various existing scalable wavelet image coding algorithms.

In chapter 4, we further extend the concept of our scalable coding algorithm, the singularity detection, to the multiwavelet transform. We estimate the Lipchitz

regularity of a signal from the interscale evolution of the individual coefficient magnitudes and the modulus of the coefficient magnitude components. We investigate the approaches for signal denoising using multiwavelet singularity detection and propose a denoising algorithm. Two selection approaches, thresholding the coefficients within the COI and thresholding the maxima coefficients inside the COI, were preliminarily applied to a 1-d test signal which is corrupted with white Guassian noise and impulsive noise. There is significant improvement compared with the WTMS denoising algorithm in the single wavelet transform.

We present the general conclusion in chapter 6, where some suggestions for further development can also be found.

Chapter 2

Preliminaries

2.1 Introduction

In this chapter, we review the background of wavelet transform and multiwavelet transform. In both parts we start with their history and development and their realization related to singularity detection. It also includes the wavelet transform modulus maxima and the wavelet transform modulus sum approaches. Then we review their applications in signal and image processing, particularly for edge detection and denoising.

2.2 Wavelet transform

In the past fifteen years, wavelet transform aroused much interest in the research area of signal processing in the engineering discipline. Wavelet transform is actually based on concepts that were developed independently under various forms in mathematics, physics and engineering. The emergence of the wavelet theory came from the connection of these multidisciplinary efforts. For signal processing, this formalization created a flow of ideas beyond the construction of new bases or transform. These ideas include the application of singularity detection for edge detection and signal and image denoising. In this part, we briefly review the history and development of the wavelet transform and wavelet singularity detection, and their applications in the signal and image processing.

2.2.1 History and development of the wavelet transform

Fourier Transform

A linear time-invariant operator L is entirely specified by the eigenvalues $\hat{h}(\omega)$,

$$\forall \omega \in \Re, \qquad Le^{i\omega t} = \hat{h}(\omega)e^{i\omega t}, \tag{2.1}$$

where the sinusoidal waves $e^{i\omega t}$ are the eigenvectors of the linear time-invariant operators. To compute Lf, a signal f is decomposed as a sum of sinusoidal

eigenvectors $\{e^{i\omega t}\}_{\omega\in\Re}$, $$f(t)=\frac{1}{2\pi}\int_{-\infty}^{+\infty}\hat{f}(\omega)e^{i\omega t}d\omega. \quad (2.2)$$

If f has finite energy, the theory of Fourier integrals proves that the amplitude $\hat{f}(\omega)$ of each sinusoidal wave $e^{i\omega t}$ is the Fourier transform of f,

$$\hat{f}(\omega)=\int_{-\infty}^{+\infty}f(t)e^{i\omega t}dt. \quad (2.3)$$

Applying the operator L to f in eqn.2.2 and inserting the eigenvector expression eqn.2.1 gives $$Lf(t)=\frac{1}{2\pi}\int_{-\infty}^{+\infty}\hat{f}(\omega)\hat{h}(\omega)e^{i\omega t}d\omega. \quad (2.4)$$

The operator L amplifies or attenuates each sinusoidal component $e^{i\omega t}$ of f by $\hat{h}(\omega)$. It is a frequency *filtering* of f.

The Fourier transform rules over linear time-invariant signal processing in early years. It proceeds simple answers to most questions in a wide range of applications such as signal transmissions or stationary signal processing. However, the Fourier transform is not tailored for transient phenomena which are in finite time segments. The Fourier coefficient is obtained in eqn.2.3 by correlating f with a sinusoidal wave $e^{i\omega t}$. Since the support of $e^{i\omega t}$ covers the whole real line, $\hat{f}(\omega)$ depends on the values $f(t)$ for all times $t\in\Re$. It becomes difficult to analyze any local property of f from $\hat{f}$. So we need local time-frequency transforms, which decompose the signal over waveforms that are well localized in time and frequency.

Time-Frequency Wedding

The uncertainty principle states that the energy spread of a function and its Fourier

transform cannot be simultaneously arbitrarily small. Motivated by quantum mechanics, in 1946 the physicist Gabor [20] defined elementary time-frequency waveforms that have a minimal spread in a time-frequency plane. To measure time-frequency information content, he proposed decomposing signals over these elementary time-frequency waveforms. By showing that such decompositions are closely related to our sensitivity to sounds, and that they exhibit important structures in speech and music recordings. Gabor demonstrated the importance of localized time-frequency signal processing.

A unified interpretation of local time-frequency decompositions follows the time-frequency energy density approach of Ville. In parallel to Gabor's contribution, in 1948 Ville [54], who was an electrical engineer, proposed analyzing the time-frequency properties of signals f with an energy density defined by

$$P_V f(t,\omega) = \int_{-\infty}^{\infty} f\left(t+\frac{\tau}{2}\right) f^*\left(t-\frac{\tau}{2}\right) e^{-i\tau\omega} d\tau .$$

Actually this distribution had already been introduced in 1932 by Wigner [55] in the context of quantum mechanics.

Windowed Fourier Transform

Gabor waveforms are constructed by translating in time and frequency a time window g:

$$g_{u,\xi}(t) = g(t-u)e^{i\xi t} .$$

The energy of $g_{u,\xi}$ is concentrated in the neighborhood of u over an interval of

size σ_t, measured by the standard deviation of $|g|^2$. Its Fourier transform is a translation by ξ of the Fourier transform $\hat{g}$ of g,

$$\hat{g}_{u,\xi}(\omega) = \hat{g}(\omega - \xi)e^{-iu(\omega-\xi)}. \tag{2.5}$$

The energy of $\hat{g}_{u,\xi}$ is therefore localized near the frequency ξ, over an interval of size σ_ω, which measures the domain where $\hat{g}(\omega)$ is non-negligible. In a time-frequency plane (t,ω), the energy spread of the $g_{u,\xi}$ is symbolically represented by the Heisenberg rectangle illustrated by figure 2.1. This rectangle is centered at (u,ξ) and has a time width σ_t and a frequency width σ_ω. The uncertainty principle proves that its area satisfies $\sigma_t \sigma_\omega \geq \frac{1}{2}$. This area is minimum when g is a Gaussian, in which case waveforms $g_{u,\xi}$ are called *Gabor* functions.

The windowed Fourier transform defined by Gabor correlates a signal f with each waveform $g_{u,\xi}$, $\quad Sf(u,\xi) = \int_{-\infty}^{\infty} f(t) g^*_{u,\xi} dt = \int_{-\infty}^{\infty} f(t) g(t-u) e^{-i\xi t} dt.$ (2.6)

It is a Fourier integral that is localized in the neighborhood of u by the window $g(t-u)$. This time integral can also be written as a frequency integral by applying the Fourier Parseval formula, $\quad Sf(u,\xi) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \hat{f}(\omega) \hat{g}^*_{u,\xi}(\omega) d\omega.$ (2.7)

The transform $Sf(u,\xi)$ thus depends only on the values $f(t)$ and $\hat{f}(\omega)$ in the time and frequency neighborhoods where the energies of $g_{u,\xi}$ and $\hat{g}_{u,\xi}$ are concentrated. Gabor interprets this as a "quantum of information" over the time-frequency rectangle illustrated in figure 2.1.

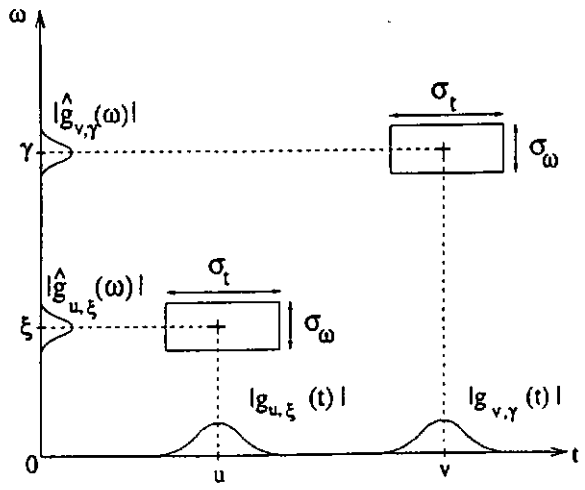


Figure 2.1 Time-frequency boxes ("Heisenberg rectangles") representing the energy spread of two Gabor atoms.

When listening to music, we perceive sound that has a frequency that varies in time. Measuring time-varying harmonics is an important application of windowed Fourier transforms in both music and speech recognition. A spectral line of f creates high amplitude windowed Fourier coefficients $Sf(u,\xi)$ at frequencies $\xi(u)$ that depend on the time u. The time evolution of such spectral components is therefore analyzed by following the location of large amplitude coefficients.

Wavelet Bases

In 1910, Haar [28] realized that one can construct a simple piecewise constant function,

$$\psi(t)=\begin{cases} 1 & \text{if } 0\le t<1/2 \\ -1 & \text{if } 1/2\le t<1 \\ 0 & \text{otherwise} \end{cases},$$

whose dilations and translations generate an orthonormal basis of $L^2(\Re)$,

$$\left\{\psi_{l,n}(t)=\frac{1}{\sqrt{2^l}}\psi\left(\frac{t-2^l n}{2^l}\right)\right\}_{(l,n)\in Z^2}. \tag{2.8}$$

Any finite energy signal $f \in L^2(\Re)$ can be decomposed over this orthogonal basis $\{\psi_{j,n}\}_{(j,n)\in Z^2}$,

$$f = \sum_{j=-\infty}^{+\infty} \sum_{n=-\infty}^{+\infty} \langle f, \psi_{j,n} \rangle \psi_{j,n} . \qquad (2.9)$$

These basis were called *wavelets*. Since $\psi(t)$ has a zero average, each partial sum

$$d_j(t) = \sum_{n=-\infty}^{+\infty} \langle f, \psi_{j,n} \rangle \psi_{j,n}(t)$$

can be interpreted as detail variations at the scale 2^j. These layers of details are added at all scales to progressively improve the approximation of f. In practical, we obtain a precise approximation by truncating the sum (eqn.2.9). The resulting approximation at a scale 2^J is $f_J(t) = \sum_{j=J}^{+\infty} d_j(t)$. The development of this piecewise constant approximation was proceeded by Daubechies and Strömberg [11,100]. The systematic theory for constructing orthonormal wavelet bases was established by Meyer and Mallat through the elaboration of *multiresolution* signal approximation [37]. The properties of orthogonal wavelets and *multiresolution* approximations did light a surprising relation with filter banks constructed with *conjugate mirror filters*.

Filter Banks

Motivated by speech compression, in 1976 Croisier, Esteban and Galand [85] introduced an invertible filter bank, which decomposes a discrete signal $f[n]$ in two signals of half its size, using a filtering and subsampling procedure. They showed that $f[n]$ can be recovered from these subsampled signals by canceling the aliasing terms

with a particular class of filters called *quadrature mirror filters*. This breakthrough led to a ten-year research effort to build a complete filter bank theory. However, besides the simple Haar filter [28], a quadrature mirror filter cannot have a finite impulse response. In 1984, Smith and Barnwell [98] and Mintzer [41] found necessary and sufficient conditions for obtaining perfect reconstruction orthogonal filters with a finite impulse response, that they called *conjugate mirror filters*. The theory was completed by the biorthogonal equations of Vetterli [52] in 1986 and the general paraunitary matrix theory of Vaidyanathan [51] in 1987. More complete presentations of filter banks can be found in the books [62,63,81,83,84].

Wavelet Transform

In reflection seismology, Morlet knew that the modulated pulses sent underground have a duration that is too long at high frequencies to separate the returns of fine, closely-spaced layers. Instead of emitting pulses of equal duration, he thus thought of sending shorter waveforms at high frequencies. Such waveforms are simply obtained by scaling a single function called a *wavelet*. Grossmann, who was working in theoretical physics, recognized Morlet's approach that was close to his own work on coherent quantum states. Nearly forty years after Gabor, Morlet and Grossmann reactivated a fundamental collaboration between theoretical physics and signal processing, which led to the formalization of the continuous wavelet transform [62].

Yet, these ideas were not totally new to mathematicians working in harmonic analysis, or to computer vision researchers studying multiscale image processing. It was thus only the beginning of gathering scientists with very different backgrounds to have rigorous discussion.

To analyze signal structures of very different sizes, it is necessary to use time-frequency waveforms with different time supports. The wavelet transform decomposes signals over dilated and translated wavelets. A wavelet ψ is a function of zero average,

$$\int_{-\infty}^{+\infty}\psi(t)dt = 0. \tag{2.10}$$

It is normalized $\|\psi\| = 1$, and centered in the neighborhood of $t = 0$. A family of time-frequency waveforms is obtained by scaling ψ by a scale parameter s, and translating it by u,

$$\psi_{u,s}(t) = \frac{1}{\sqrt{s}}\psi\left(\frac{t-u}{s}\right). \tag{2.11}$$

The wavelet transform of f at the scale s and position u is computed by correlating f with a wavelet function

$$Wf(u,s) = \langle f, \psi_{u,s}\rangle = \int_{-\infty}^{+\infty} f(t)\frac{1}{\sqrt{s}}\psi^*\left(\frac{t-u}{s}\right)dt = f * \overline{\psi}_s(u) \tag{2.12}$$

with $$\overline{\psi}_s(t) = \frac{1}{\sqrt{s}}\psi^*\left(\frac{-t}{s}\right).$$

The Fourier transform of $\overline{\psi}_s(t)$ is $\hat{\overline{\psi}}_s(\omega) = \sqrt{s}\hat{\psi}^*(s\omega)$. Since $\hat{\psi}(0) = \int_{-\infty}^{+\infty}\psi(t)dt = 0$, it appears that $\hat{\psi}$ is the transfer function of a band-pass filter. The convolution (eqn.2.12) computes the wavelet transform with dilated band-pass filters.

Wavelet Transform and Filter Banks

The equivalence between the continuous time wavelet theory and discrete filter banks led to the interface between digital signal processing and harmonic analysis. However, dilations are not defined over discrete sequences and discrete wavelet bases have therefore a more complicated structure. The regularity of a discrete sequence is also not well defined, and this makes it more difficult to interpret the amplitudes of wavelet coefficients. So the role of the theory of continuous time functions is to give asymptotic results for discrete sequences with sampling intervals tending to zero. It is because these asymptotic results are precise enough to understand the behaviour of discrete algorithm.

Frame Theory

The continuous windowed Fourier transform $Sf(u,\xi)$ and the wavelet transform $Wf(u,s)$ are 2-d representations of a 1-d signal f. This indicates the existence of some redundancy that can be reduced and even removed by subsampling the parameters of these transforms. Eqn.2.6 and eqn.2.12 can be written as inner products in $L^2(\Re)$,

$$Sf(u,s) = \int_{-\infty}^{\infty} f(t) g^*_{u,\xi}(t)\,dt = \langle f, g_{u,\xi} \rangle$$

and

$$Wf(u,s) = \int_{-\infty}^{\infty} f(t) \psi^*_{u,s}(t)\,dt = \langle f, \psi_{u,s} \rangle .$$

Subsampling both transforms defines a complete signal representation if any signal can be reconstructed from linear combinations of discrete families of windowed

$\{g_{u_n,\xi_k}\}_{(n,k)\in Z^2}$ and $\{\psi_{u_n,s_j}\}_{(j,n)\in Z^2}$. The frame theory discusses what conditions these families of waveforms must meet if they are to provide stable and complete representations. Signal reconstructions from regular and irregular samplings are examples of applications. The frame theory was originally developed by Duffin and Schaeffer [19] to reconstruct band-limited signals f from irregularly spaced samples $\{f(t_n)\}_{n\in Z}$. The general sampling theorem, which proves that if f has a Fourier transform included in $[-\pi/T,\pi/T]$, then $f(t_n)=\frac{1}{T}\langle f(t),h_T(t-t_n)\rangle$ with $h_T(t)=\frac{\sin(\pi t/T)}{\pi t/T}$, gives a sufficient condition for reconstructing a signal from its samples. Daubechies [12,67] proves several necessary and sufficient conditions for constructing a bounded frame of $L^2(\Re)$ with dilated windows, windowed Fourier transform and wavelet transform.

2.2.2 Important features of the wavelet transform

Like a windowed Fourier transform, a wavelet transform can measure the time evolution of frequency transients. To analyze the phase information of signals, a complex analytic wavelet is used. We skip the details of the properties of the analytic wavelet transform and its application to the measurement of instantaneous frequencies here. It is because we are interested in sharp signal transitions, which are detected by real wavelets. So we introduce the elementary properties of real wavelets as the

following.

Suppose ψ is real. Since it has a zero average, a wavelet coefficient $Wf(u,s)$ measures the variation of f in the neighborhood of u whose size is proportional to s. When the scale s goes to zero, the decay of the wavelet coefficients characterizes the regularity of f in the neighborhood of u. Sharp signal transitions create high amplitude wavelet coefficients. In section 2.2.4, the pointwise regularity of f is related to the asymptotic decay of the wavelet transform $Wf(u,s)$ when s goes to zero. Singularities are detected by following across scales the local maxima of the wavelet transform. If $Wf(u,s)$ has no modulus maxima at fine scales, then f is locally regular. In images, high amplitude wavelet coefficients indicate the position of edges, which are sharp variations of image intensity. Different scales provide the contours of image structures of varying sizes. Such multiscale edge detection is particularly effective for pattern recognition in computer vision [4].

A real wavelet transform is complete and maintains energy conservation, as long as the wavelet satisfies a weak *admissibility condition*. This theorem was first proved in 1964 by the mathematician Calderón [3]. Calderón defines a wavelet transform as a convolution operator that decomposes the identity. Grossmann and Morlet [4] also proved the same formula for signal processing. If $\hat{\psi}(0)=0$ and $\hat{\psi}(\omega)$ is continuously differentiable then the admissibility condition is satisfied. With the

admissibility condition satisfied, one can verify that $\hat{\psi}(\omega)$ is continuously differentiable if ψ has a sufficient time decay $\int_{-\infty}^{+\infty}(1+|t|)|\psi(t)|dt < +\infty$.

Multiresolution Approximation

Orthogonal wavelets (eqn.2.8) dilated by 2^j carry signal variations at the resolution 2^{-j}. The construction of these bases can thus be begun with multiresolution signal approximations. This leads to equivalence between wavelet bases and conjugate mirror filters used in discrete multirate filter banks. The multiresolution theory of orthogonal wavelets [37] proves that any conjugate mirror filter characterizes a wavelet ψ that generates an orthonormal basis of $L^2(\Re)$. Eqn.2.9 can indeed be interpreted as the difference between two approximations of f at the resolution 2^{-j+1} and 2^{-j}. Multiresolution approximations compute the approximation of signals at various resolutions with orthogonal projections on different spaces $\{V_j\}_{j\in Z}$

Definition 2.1 ***Multiresolution***

A sequence $\{V_j\}_{j\in Z}$ *of closed subspaces of* $L^2(\Re)$ *is a multiresolution approximation if the following 6 properties are satisfied:*

$$\forall (j,k)\in Z^2, \qquad f(t)\in V_j \Leftrightarrow f(t-2^j k)\in V_j, \tag{2.13}$$

$$\forall j \in Z, \qquad V_{j+1}\subset V_j, \tag{2.14}$$

$$\forall j \in Z, \qquad f(t)\in V_j \Leftrightarrow f\left(\frac{t}{2}\right)\in V_{j+1}. \tag{2.15}$$

$$\lim_{j\to+\infty} V_j = \bigcap_{j=-\infty}^{+\infty} V_j = \{0\}, \tag{2.16}$$

$$\lim_{j\to-\infty} V_j = closure\left(\bigcup_{j=-\infty}^{+\infty} V_j\right) = L^2(R). \tag{2.17}$$

There exists θ *such that* $\{\theta(t-n)\}_{n\in Z}$ *is a Riesz basis of* V_0.

Property eqn.2.13 means that V_j is invariant by any translation proportional to the scale 2^j, which characterizes the signal approximation at the resolution 2^{-j}. Eqn.2.14 is a causality property which proves that an approximation at a resolution 2^{-j} contains all the necessary information to compute an approximation at a coarser resolution 2^{-j-1}. Dilating functions in V_j by 2 enlarges the details by 2 and eqn.2.15 guarantees that it defines an approximation at a coarser resolution 2^{-j-1}. When the resolution 2^{-j} goes to zero eqn.2.16 implies that we lose all the details of f and $\lim_{j\to+\infty}\left\| P_{V_j} f \right\| = 0$. On the other hand, when the resolution 2^{-j} goes to $+\infty$, property eqn.2.17 imposes that the signal approximation converges to the original signal and $\lim_{j\to+\infty}\left\| f - P_{V_j} f \right\| = 0$. When the resolution 2^{-j} increases, the decay rate of the approximation error $\left\| f - P_{V_j} f \right\|$ depends on the regularity of f. Section 2.2.4 relates this error to the uniform Lipschitz regularity of f.

2.2.3 Realization of the wavelet transform

Frame Construction

In the realization of the wavelet transform, we practically seek for the minimal

requirement for a complete signal representation. Frame theory reveals the possibility of removing the redundancy of wavelet transform representation by subsampling the transform under various conditions. The frame theory analyzes the completeness, stability and redundancy of linear discrete signal representations. A frame is a family of vectors $\{\phi_n\}_{n\in\Gamma}$ that characterizes any signal f from its inner products $\{\langle f,\phi_n\rangle\}_{n\in\Gamma}$. Intuitively, to construct a frame we need to cover the time-frequency plane with the Heisenberg boxes of the corresponding discrete wavelet family. A wavelet $\psi_{u,s}$ has an energy in time that is centered at u over a domain proportional to s. Over positive frequencies, its Fourier transform $\hat{\psi}_{u,s}$ has a support centered at a frequency η/s, with a spread proportional to $1/s$. To obtain a full cover, we sample s along an exponential sequence $\{a^j\}_{j\in Z}$, with a sufficiently small dilation step $a>1$. The time translation u is sampled uniformly at intervals proportional to the scale a^j. Let us denote a wavelet frame $\psi_{j,n}(t)=\frac{1}{\sqrt{a^j}}\psi\left(\frac{t-nu_o a^j}{a^j}\right)$ yields inner products that sample the continuous wavelet transform at time intervals $a^j u_o$,

$$\langle f,\psi_{j,n}\rangle = f*\overline{\psi}_{a^j}(na^j u_o)=Wf(na^j u_o,a^j).$$

Translating f by τ gives

$$\langle f_\tau,\psi_{j,n}\rangle = f*\overline{\psi}_{a^j}(na^j u_o-\tau)=Wf(na^j u_o-\tau,a^j).$$

We can see that if the sampling interval $a^j u_o$ is large relative to the rate of variation of $f*\overline{\psi}_{a^j}(t)$, then the coefficients $\langle f,\psi_{j,n}\rangle$ and $\langle f_\tau,\psi_{j,n}\rangle$ may take very different

values that are not translated with respect to one another. This is particularly acute for wavelet orthogonal bases where u_o is maximum. The same translation distortion phenomena also appear in windowed Fourier frames.

Fast Dyadic Transform

In pattern recognition, it is important to construct signal representations that are translation invariant. When a pattern is translated, its numerical descriptors should be translated but not modified. Indeed, a pattern search is particularly difficult if its representation depends on its location. Continuous wavelet transforms and windowed Fourier transform provide translation-invariant representations, but uniformly sampling the translation parameter destroys this translation invariance. There are several strategies for maintaining the translation invariance of a wavelet transform. If the sampling interval $a^j u_o$ is small enough then the samples of $f * \bar{\psi}_{a^j}(t)$ are approximately translated when f is shifted. The dyadic wavelet transform is a translation-invariant representation that discretizes the scale s into a dyadic sequence $\{2^j\}_{j\in Z}$ but does not sample the translation factor u. It thus creates a highly redundant signal representation. Suppose that the scaling functions and wavelets ϕ, ψ, $\tilde{\phi}$ and $\tilde{\psi}$ are designed with the filters h, g, $\tilde{h}$ and $\tilde{g}$. A fast dyadic wavelet transform is calculated with a filter bank algorithm called in French the *algorithme à trous*, introduced by Holscheneider, Kronland-Martinet, Morlet and

Tchamitchian [71]. Consider the signal $f(t) = f(N^{-1}t)$, whose samples have distance equal to 1. A change of variable in the dyadic wavelet transform integral shows that $Wf(u,2^j) = N^{-1/2} Wf(Nu, N2^j)$.

We suppose that the samples $a_o[n]$ of the input discrete signal are not equal to $f(n)$ but to a local average of f in the neighborhood of $t = n$. The samples $a_o[n]$ are written as averages of $f(t)$ weighted by the scaling kernels $\phi(t-n)$,

$$a_o[n] = \langle f(t), \phi(t-n) \rangle = \int_{-\infty}^{+\infty} f(t)\phi(t-n)dt .$$

For any $j \geq 0$, we denote

$$a_j[n] = \langle f(t), \phi_{2^j}(t-n) \rangle \quad \text{with} \quad \phi_{2^j}(t) = \frac{1}{\sqrt{2^j}} \phi\left(\frac{t}{2^j}\right).$$

For $j > 0$, $\quad d_j[n] = Wf(n,2^j) = \langle f(t), \psi_{2^j}(t-n) \rangle .$

Note that we insert $2^j - 1$ zeros between each sample of the filters. Inserting zeros in the filters creates holes (*trous* in French). Let $\bar{x}_j[n] = x_j[-n]$. The following proposition gives convolution formulas that are cascaded to compute a dyadic wavelet transform and its inverse.

Proposition 2.1 ***Fast Dyadic Transform***

For any $j \geq 0$, $\quad a_{j+1}[n] = a_j * \bar{h}_j[n], \qquad d_{j+1}[n] = a_j * \bar{g}_j[n],$

and $\quad a_j[n] = \frac{1}{2}\left(a_{j+1} * \tilde{h}_j[n] + d_{j+1} * \tilde{g}_j[n]\right).$

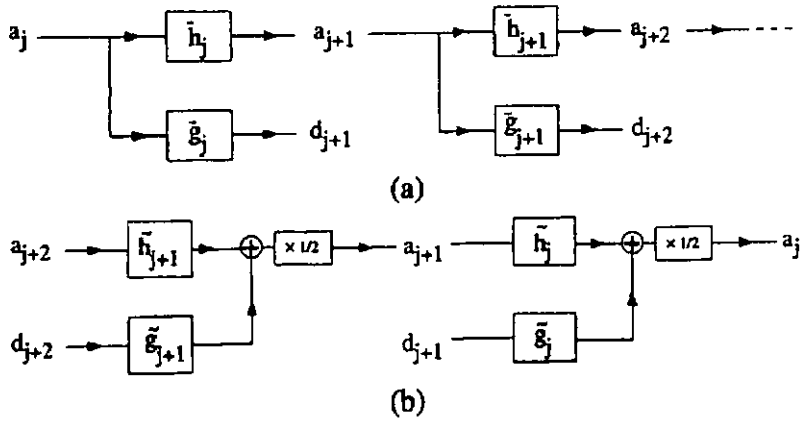


Figure 2.2 Fast dyadic transform. (a) The dyadic wavelet coefficients are computed by cascading convolutions with dilated filters. (b) The original signal is reconstructed through convolutions.

To reduce the representation size while maintaining translation invariance, one can use an adaptive sampling scheme, where the sampling grid is automatically translated when the signal is translated. For each scale a^j, $Wf(u,a^j) = f * \overline{\psi}_{a^j}(u)$ can be sampled at location u where $|Wf(u,a^j)|$ is locally maximum. The resulting representation is translation invariant since the local maxima positions are translated when f and hence $f * \overline{\psi}_{a^j}$ are translated. However, the reconstruction is complicated. To study the completeness and stability of wavelet maxima representations, Mallat and Zhong introduced an alternate projection algorithm [39] that recovers signal approximations from their wavelet maxima. Several other algorithms have been proposed more recently [10,27,64]. For general dyadic wavelets, Meyer [77] and Berman [1] proved that exact reconstruction is not possible, and signals with the same wavelet maxima differ from each other only slightly, which

explains the success of numerical reconstructions [39].

Filter Bank Implementation and Perfect Reconstruction

It was proved that multiresolution approximations are entirely characterized by a particular discrete filter that governs the loss of information across resolutions. These discrete filters provide a *simple* procedure for designing and synthesizing orthogonal wavelet bases. Let ϕ be a scaling function that generates an orthogonal basis of each space V_j. We study the properties of ϕ which guarantee that the spaces V_j satisfy all conditions of a multiresolution approximation. It was proved that any scaling function is specified by a discrete filter called a *conjugate mirror filter.*

The multiresolution causality property eqn.2.14 imposes that $V_j \subset V_{j-1}$. In particular $2^{-1/2}\phi(t/2) \in V_1 \subset V_o$. Since $\{\phi(t-n)\}_{n\in Z}$ is an orthonormal basis of V_o, we can decompose

$$\frac{1}{\sqrt{2}}\phi\left(\frac{t}{2}\right) = \sum_{n=-\infty}^{+\infty} h[n]\phi(t-n), \tag{2.18}$$

with

$$h[n] = \left\langle \frac{1}{\sqrt{2}}\phi\left(\frac{t}{2}\right), \phi(t-n) \right\rangle. \tag{2.19}$$

This scaling equation relates a dilation of ϕ by 2 to its integer translations. The sequence $h[n]$ will be interpreted as a discrete filter. Since $\{\phi_{j,n}\}_{n\in Z}$ and $\{\psi_{j,n}\}_{n\in Z}$ are orthonormal bases of V_j and W_j ($V_{j-1} = V_j \oplus W_j$), the projection in these spaces is characterized by $a_j[n] = \langle f, \phi_{j,n} \rangle$ and $d_j[n] = \langle f, \psi_{j,n} \rangle$.

The following theorem [38,95] shows that these coefficients are calculated with a

cascade of discrete convolutions and subsamplings. We denote $\bar{x}[n]=x[-n]$ and

$$\breve{x}=\begin{cases} x[p] & \text{if } n=2p \\ 0 & \text{if } n=2p+1 \end{cases}. \tag{2.20}$$

Theorem 2.1 ***Fast Orthogonal Wavelet Transform***

At the decomposition, $$a_{j+1}[p]=\sum_{n=-\infty}^{+\infty} h[n-2p]a_j[n]=a_j * \bar{h}[2p] \tag{2.21}$$

$$d_{j+1}[p]=\sum_{n=-\infty}^{+\infty} g[n-2p]d_j[n]=d_j * \bar{g}[2p]. \tag{2.22}$$

At the reconstruction, $$a_j[p]=\sum_{n=-\infty}^{+\infty} h[p-2n]a_{j+1}[n]+\sum_{n=-\infty}^{+\infty} g[p-2n]d_{j+1}[n]$$

$$=\breve{a}_{j+1} * h[p]+\breve{d}_{j+1} * g[p]. \tag{2.23}$$

Theorem 2.1 proves that a_{j+1} and d_{j+1} are computed by taking every other sample of the convolution of a_j with $\bar{h}$ and $\bar{g}$ respectively, as illustrated by figure 2.3. The filter $\bar{h}$ removes the higher frequencies of the inner product sequence a_j whereas $\bar{g}$ is a high-pass filter which collects the remaining highest frequencies. The reconstruction eqn.2.23 is an interpolation that inserts zeros to expand a_{j+1} and d_{j+1} and filter these signals, as shown in figure 2.3. An *orthogonal wavelet representation* $a_L=\langle f,\phi_{L,n}\rangle$ is composed of wavelet coefficients of f at scales $2^L<2^j\le 2^J$ plus the remaining approximation at the largest scale 2^J,

$$\{d_j\}_{L<j\le J},\quad a_J$$

It is computed from a_L by iterating eqn.2.21 and 2.22 for $L\le j<J$. The original signal a_L is recovered from this wavelet representation by iterating the reconstruction eqn.2.23 for $J>j\ge L$.

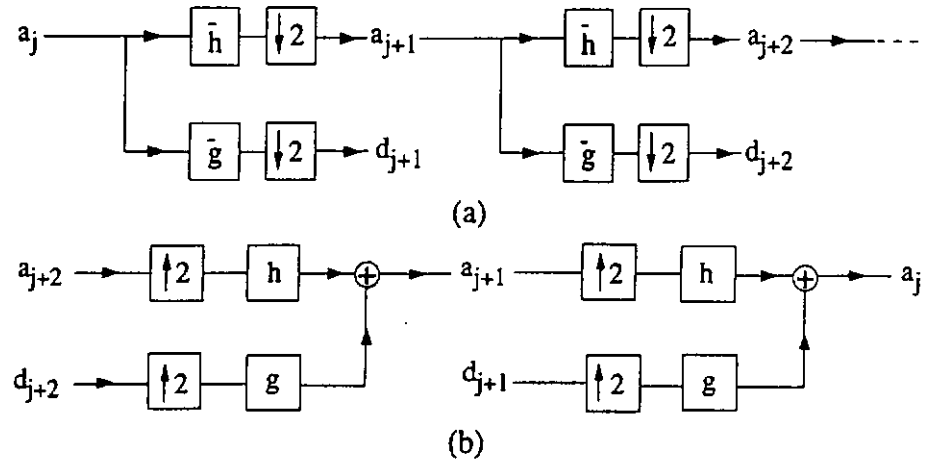


Figure 2.3 Fast wavelet transform. (a) The forward is computed with a cascade of filterings followed by a factor 2 subsampling. (b) The inverse reconstructs progressively by inserting zeros between the samples, filtering and adding the output.

A two-channel multirate filter bank convolves a signal a_o with a low-pass filter $\bar{h}[n]=h[-n]$ and a high-pass filter $\bar{g}[n]=g[-n]$, and subsamples by 2 the output:

$$a_1[n]=a_o * \bar{h}[2n] \quad \text{and} \quad d_1[n]=a_o * \bar{g}[2n]. \tag{2.24}$$

A reconstructed signal $\tilde{a}_o$ is obtained by filtering the zero expanded signals with a dual low-pass filter $\tilde{h}$ and a dual high-pass filter $\tilde{g}$, as shown in figure 2.4. With the zero insertion (eqn.2.20) it yields

$$\tilde{a}_o[n]=\check{a}_1 * \tilde{h}[n]+\check{d}_1 * \tilde{g}[n]. \tag{2.25}$$

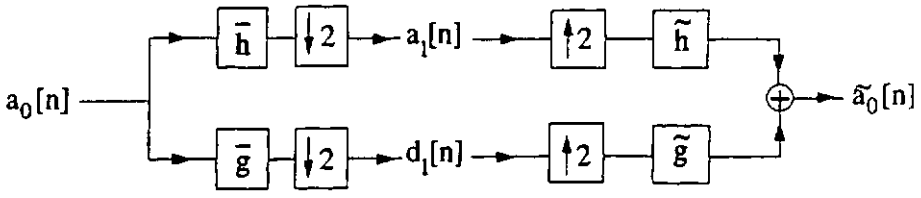


Figure 2.4 A two-channel multirate filter bank. The input signal is filtered by a low-pass and a high-pass filter and subsampled. The reconstruction is performed by inserting zeros and filtering with dual filters $\tilde{h}$ and $\tilde{g}$.

The following theorem [52] gives the necessary and sufficient conditions on h, g, $\bar{h}$ and $\bar{g}$ to guarantee a perfect reconstruction $\tilde{a}_o = a_o$.

Theorem 2.2 ***Perfect Reconstruction for Two-channel Filter Bank***

The filter bank performs an exact reconstruction for any input signal if and only if

$$\hat{h}^*(\omega+\pi)\hat{\tilde{h}}(\omega)+\hat{g}^*(\omega+\pi)\hat{\tilde{g}}(\omega)=0 \tag{2.26}$$

and

$$\hat{h}^*(\omega)\hat{\tilde{h}}(\omega)+\hat{g}^*(\omega)\hat{\tilde{g}}(\omega)=2. \tag{2.27}$$

Theorem 2.2 proves that the reconstruction filters $\tilde{h}$ and $\tilde{g}$ are entirely specified by the decomposition filters h and g. In matrix form, it can be rewritten as

$$\begin{pmatrix} \hat{h}(\omega) & \hat{g}(\omega) \\ \hat{h}(\omega+\pi) & \hat{g}(\omega+\pi) \end{pmatrix} \times \begin{pmatrix} \hat{\tilde{h}}^*(\omega) \\ \hat{\tilde{g}}^*(\omega) \end{pmatrix} = \begin{pmatrix} 2 \\ 0 \end{pmatrix} \tag{2.28}$$

The inversion of this 2×2 matrix yields

$$\begin{pmatrix} \hat{\tilde{h}}^*(\omega) \\ \hat{\tilde{g}}^*(\omega) \end{pmatrix} = \frac{2}{\Delta(\omega)} \begin{pmatrix} \hat{g}(\omega+\pi) \\ -\hat{h}(\omega+\pi) \end{pmatrix} \tag{2.29}$$

where $\Delta(\omega)$ is the determinant

$$\Delta(\omega)=\hat{h}(\omega)\hat{g}(\omega+\pi)-\hat{h}(\omega+\pi)\hat{g}(\omega). \tag{2.30}$$

The reconstruction filters are stable only if the determinant does not vanish for all $\omega \in [-\pi,\pi]$. This result is extended to multirate filter banks with an arbitrary number M channels by showing that the resulting matrices of filters satisfy paraunitary properties in [51].

Boundary Problems

For the realization of the wavelet transform, we need to handle the filtering at the boundaries of finite signals. *Boundary* wavelets are designed in order to provide this necessary complement. The main difficulty of constructing boundary wavelets is to keep their vanishing moments. There are three common approaches to solve this problem, periodic wavelets, folded wavelets and custom-designed boundary wavelets. In the first approach, we treat $f[n]$ and the wavelets $\psi_j[n]$ as periodic signals of period N. The discrete wavelet transform can then be written as a circular convolution which is calculated with the fast Fourier transform algorithm requiring $O(N\log_2 N)$ operations:

$$Wf[n,a^j]=\sum_{m=0}^{N-1} f[m]\psi_j^*[m-n] \tag{2.31}$$

where $\overline{\psi}_j[n]=\psi_j^*[-n]$. If $a=2^{1/v}$, there are $v\log_2(N/(2K))$ scales $a^j\in[2N^{-1},K^{-1}]$. The total number of operations to compute the wavelet transform over all scales is therefore $O\left(vN(\log_2 N)^2\right)$ [44]. Periodic wavelets have no vanishing moments at the boundary, whereas folded wavelets, which is done by symmetric extension of the signal at the boundaries, have only one vanishing moments. The custom-designed boundary wavelets [9] can have many vanishing moments but are more complicated to construct.

2.2.4 Wavelet singularity detection

Lipschitz Regularity

The Taylor formula relates the differentiability of a signal $f \in L^2(\Re, C)$ to local polynomial approximations. Suppose that f is m times differentiable in $[v-h, v+h]$. Let p_v be the Taylor polynomial in the neighborhood of v:

$$p_v(t) = \sum_{k=0}^{m-1} \frac{f^{(k)}(v)}{k!}(t-v)^k .$$

The Taylor formula proves that the approximation error

$$\varepsilon_v(t) = f(t) - p_v(t)$$

satisfies $\forall t \in [v-h, v+h], \quad |\varepsilon_v(t)| \le \frac{|t-v|^m}{m!} \sup_{u \in [v-h, v+h]} |f^m(u)|.$

The m^{th} order differentiability of f in the neighborhood of v yields an upper bound on the error $\varepsilon_v(t)$ when t tends to v. The Lipschitz regularity refines this upper bound with non-integer exponents. Lipschitz exponents are also called Hölder exponents in the mathematical literature.

Definition 2.2 ***Lipschitz regularity***

A function f is pointwise Lipschitz $\alpha \ge 0$ at v, if there exist $K > 0$, and a polynomial p_v of degree $m = \lfloor \alpha \rfloor$ such that

$$\forall v \in \Re, \quad |f(t) - p_v(t)| \le K|t-v|^\alpha . \tag{2.32}$$

A function f is uniformly Lipschitz α over $[a,b]$ if it satisfies the above

condition for all $v \in [a,b]$, *with a constant* K *that is independent of* v. *The Lipschitz regularity of* f *at* v *or over* $[a,b]$ *is the sup of the* α *such that* f *is Lipschitz* α.

Remarks

1. *At each* v *the polynomial* $p_v(x)$ *is uniquely defined. If* f *is* $m = \lfloor \alpha \rfloor$ *times continuously differentiable in the neighborhood of* v, *then* p_v *is the Taylor expansion of* f *at* v.

2. *Since pointwise Lipschitz exponents may vary arbitrary from abscissa to abscissa, to obtain a more global measurement of regularity, that is,* f *has the same Lipschitz regularity over the neighborhood of* v, *we need uniform Lipschitz regularity.*

3. *If* $0 \le \alpha < 1$, *then there exists* $K > 0$, *and a polynomial* $p_v(x) = f(v)$, *such that the Lipschitz condition in definition 2.1 becomes*

$$\forall t \in \Re, \ |f(t) - f(v)| \le K|t - v|^{\alpha}.$$

4. *A function that is bounded but discontinuous at* v *is Lipschitz* 0 *at* v.

5. *If the Lipschitz regularity is* $\alpha < 1$ *at* v, *then* f *is not differentiable at* v *and* α *characterizes the singularity type.*

Estimation of the Lipschitz Regularity

To analyze signal structures of very different sizes, it is necessary to use

time-frequency waveforms with different time supports. The wavelet transform decomposes signals over dilated and translated wavelet. A wavelet is a function $\psi \in L^2(\Re)$ with a zero average, $\int_{-\infty}^{+\infty} \psi(t)dt = 0$.

It is normalized $\|\psi\| = 1$, and centered in the neighborhood of $t = 0$.

Definition 2.3 ***Wavelet Transform***

Let ϕ be a scaling function and ψ be the corresponding wavelet generating a wavelet orthonormal basis of $L^2(\Re)$. Let $\phi_{u,s}(x) = \frac{1}{\sqrt{s}}\phi\left(\frac{t-u}{s}\right)$ be the set of translation by u and dilation by s of the mother wavelet ϕ. Let $\psi_{u,s}(t) = \frac{1}{\sqrt{s}}\psi\left(\frac{t-u}{s}\right)$ be the set of translation by u and dilation by s of the mother wavelet ψ, where $\|\psi_{u,s}\| = 1$. The wavelet transform of $f \in L^2(\Re, C)$ at time u and scale s is $Wf(u,s) = \langle f, \psi_{u,s} \rangle = \int_{-\infty}^{\infty} f(t)\frac{1}{\sqrt{s}}\psi\left(\frac{t-u}{s}\right)dt$,

and $Lf(u,s) = \langle f, \phi_{u,s} \rangle$.

The uniform Lipschitz regularity of f over $\Re$ can be related to the asymptotic decay of its Fourier transform. The Fourier transform is a powerful tool for measuring the minimum global regularity of functions. However, it is not possible to analyze the regularity of f at a particular point v from the decay of $|\hat{f}(\omega)|$ at high frequency ω. In contrast, since wavelets are well localized in time, the wavelet transform gives Lipschitz regularity over intervals and at points.

To measure the local regularity of a signal, it is not so important to use a wavelet

with a narrow frequency support. Vanishing moments are crucial. If the wavelet has n vanishing moments then it can be shown that the wavelet transform can be interpreted as a multiscale differential operator of order n. This yields a first relation between the differentiability of f and its wavelet transform decay at fine scales.

The Lipschitz property eqn.2.32 approximates f with a polynomial p_v in the neighborhood of v, $f(t)=p_v(t)+\varepsilon_v(t)$ with $|\varepsilon_v(t)|\le K|t-v|^{\alpha}$.

A wavelet transform estimates the exponent α by ignoring the polynomial p_v. For this purpose, we use a wavelet that has $n>\alpha$ vanishing moments,

$$\int_{-\infty}^{\infty} t^k \psi(t)dt = 0 \quad \text{for} \quad 0 \le k < n.$$

A wavelet with n vanishing moments is orthogonal to polynomials of degree $n-1$. Since $\alpha<n$, the polynomial p_v has degree at most $n-1$. With the change of variable $t'=(t-u)/s$,

$$Wp_v(u,s)=\langle p_v,\psi_{u,s}\rangle=0. \tag{2.33}$$

Since $f=p_v+\varepsilon_v$, $Wf(u,s)=W\varepsilon_v(u,s)$.

In this part we explain how to measure α from $|Wf(u,s)|$ when u is in the neighborhood of v. In fact, the decay of the wavelet transform amplitude across scales is related to the uniform and pointwise Lipschitz regularity of the signal. Measuring this asymptotic decay is equivalent to zooming into signal structures with a scale that goes to zero. We suppose that the wavelet ψ has n vanishing moments

and is n times continuously differentiable with derivatives that have a fast decay. This means that for any $0 \le k \le n$ and $m \in \mathrm{N}$ there exists C_m such that

$$\forall t \in \Re, \qquad \left|\psi^{(k)}(t)\right| \le \frac{C_m}{1+|t|^m}. \tag{2.34}$$

The following theorems relate the uniform and pointwise Lipschitz regularity of f on an interval to the amplitude of its wavelet transform at fine scales. All theorems remain valid if we restrict s to the dyadic scales $\{2^j\}_{j\in Z}$.

Theorem 2.3 ***Uniform Lipschitz Regularity***

If $f \in L^2(\Re, C)$ *is uniformly Lipschitz* $\alpha \le n$ *over* $[a,b]$, *then there exists* $A > 0$ *such that* $\forall (u,s) \in [a,b] \times \Re^+, \quad |Wf(u,s)| \le As^{\alpha+1/2}$.

Conversely, suppose that f *is bounded and that* $Wf(u,s)$ *satisfies the above condition for an* $\alpha \le n$ *that is not an integer. Then* f *is uniformly Lipschitz* α *on* $[a+\varepsilon, b+\varepsilon]$, *for any* $\varepsilon > 0$.

The proof of theorem 2.3 can be found in [40,74,75,76].

The study of pointwise Lipschitz exponents with the wavelet transform is a delicate topic long time ago. Characterizing the regularity of f at a point v can be difficult because f may have very different types of singularities that are aggregated in the neighborhood of v. The following theorem gives a necessary condition and a sufficient condition on the wavelet transform for estimating the Lipschitz regularity of f at a point v. We still assume the wavelet ψ has n derivatives and a fast decay.

Theorem 2.4 ***Pointwise Lipschitz Regularity***

If $f \in L^2(\Re, C)$ *is pointwise Lipschitz* $\alpha \le n$ *at* v, *then there exists* $A > 0$, *such that*

$$\forall (u,s) \in \Re \times \Re^+, \quad |Wf(u,s)| \le A s^{\alpha+1/2}\left(1+\left|\frac{u-v}{s}\right|^{\alpha}\right).$$

Conversely, if $\alpha < n$ *is not an integer and there exist* A *and* $\alpha' < \alpha$ *such that*

$$\forall (u,s) \in \Re \times \Re^+, \quad |Wf(u,s)| \le A s^{\alpha+1/2}\left(1+\left|\frac{u-v}{s}\right|^{\alpha'}\right),$$

then f *is Lipschitz* α *at* v.

The proof of theorem 2.4 can also be found in [40,74,75,76].

Theorem 2.3 and 2.4 prove that the local Lipschitz regularity of f at v depends on the decay at fine scales of $|Wf(u,s)|$ in the neighborhood of v. Measuring this decay directly in the time-scale plane (u,s) is not necessary. Singularities create sequences of maxima that converge towards the corresponding location at fine scales, and the Lipschitz regularity can be calculated from the decay of the maxima amplitude. We use the term *modulus maximum* to describe any point (u_o, s_o) such that $|Wf(u,s)|$ is locally maximum at $u = u_o$. This implies that $\frac{\partial Wf(u_o, s_o)}{\partial u} = 0$. The local maximum should be a strict local maxima in either the right or the left neighborhood of u_o, to avoid having any local maxima when $|Wf(u,s)|$ is constant. Singularities are detected by finding the abscissa by following the wavelet modulus maxima which converge at fine scales. We call maxima line any

connected curve $s(u)$ in the scale-space plane (u,s) along which all points are modulus maxima.

It was observed that a wavelet transform may have a sequence of local maxima that converge to an abscissa v even though f is perfectly regular at v. To detect singularities it is therefore not sufficient to follow the wavelet modulus maxima across scales. The Lipschitz regularity is calculated from the decay of the modulus maxima amplitude. Let us suppose that for $s < s_0$ all modulus maxima that converge to v are including in a cone $|u-v| \le Ks$, where K is the size of the support of the wavelet. f does not have oscillations that accelerate in the neighborhood of v. That is, when potential singularity at v is isolated, the decay of $|Wf(u,s)|$ in the neighborhood of v is controlled by the decay of the modulus maxima included in the cone $|u-v| \le Ks$. Recall theorem 2.3 and take log at both sides,

$$\log_2|Wf(u,s)| \le \log_2 A + (\alpha + 1/2)\log_2 s .$$

The Lipschitz regularity at v is thus the maximum slope of $\log_2|Wf(u,s)|$ is a function of $\log_2 s$ along the maxima lines converging to v.

In numerical calculation, the finest scale of the wavelet transform is limited by the resolution of the discrete data, $s \ge \lambda N^{-1}$. λ must be large enough to avoid sampling coarsely the wavelets at the finest scale. The α of a singularity is then estimated by measuring the decay slope of $\log_2|Wf(u,s)|$ as a function of $\log_2 s$ for

$2^J \geq s \geq \lambda N^{-1}$. We usually take the largest scale 2^J to be smaller than the distance between two consecutive singularities to avoid having other singularities influence the value of $Wf(u,s)$.

Later it was found that there are many errors and ambiguities in tracing the maxima curves in the scale-space plane. So the wavelet transform modulus sum (WTMS) was proposed [32]. The selection of wavelet coefficients by the interscale ratio and difference of WTMS inside the COI was used to select wavelet coefficients, instead of directly determine the Lipschitz regularity of a signal. Indeed, the main difference of using WTMS instead of WTMM is that the processing of WTMS is over the regularity located wavelet coefficients while the processing of WTMM is over irregularly located maximum points of the wavelet coefficients. This significantly decreases the computational complexity in processing.

Definition 2.4 ***Wavelet Transform Modulus Sum***

Suppose the wavelets have compact support $[-K,K]$. *The cone of influence is* $COI_{j,v} = \{u : |u-v| \leq K(2^j)\}$ *for any particular point* v. *The integral sum of the magnitude of the wavelet coefficients inside this COI is determined as*

$$N_{j,v}f = \int_{COI_{j,v}} |W_j f(u)| du .$$

Theorem 2.5

If $f \in L^2(\Re, C)$ *is uniformly Lipschitz* α *in the neighborhood of* v, *if and only if there exists a finite constant* $A > 0$ *such that* $N_{j,v} f \le A\left(2^j\right)^{\alpha+1}$.

This theorem is proved in [73].

Again we need not estimating the Lipschitz exponent directly. It is known that the Lipschitz exponent α of Guassian noise usually possesses negative value when it is measured by the wavelet transform with wavelet of one vanishing moment [40]. In the case $\alpha \le -1$, $N_{j+1,v} f \le N_{j,v} f$, for $1 \le j < J$,

where J is the total number of scales. It implies that for a strong irregular point which has $\alpha \le -1$, we can easily detect it by measuring the interscale ratio of $N_{j,v} f$ such that $\frac{N_{j+1,v} f}{N_{j,v} f} = 2^{\alpha+1} \le 1$ for $1 \le j < J$ and $\alpha \le -1$.

That is, the function $N_{j,v} f$ will decrease or remain the same as the scale increases. If the point v corresponds to the edge or the regular part of the function f, then $\alpha \ge 0$. This point can also be easily detected by measuring the interscale ratio of $N_{j,v} f$ such that $\frac{N_{j+1,v} f}{N_{j,v} f} = 2^{\alpha+1} \ge 2$ for $1 \le j < J$ and $\alpha \ge 0$.

That is, the function $N_{j,v} f$ will increase at least twice as the scale increases. By selecting the wavelet coefficients fulfilling the above condition, we can effectively distinguish and remove noise while the edges and the regular parts can be preserved.

Since it is observed in [32] that some small irregular signals, which have $-1<\alpha<0$, will have their wavelet coefficients wrongly fulfill the above criterion, another criterion, the interscale difference of WTMS, was introduced,

$$N_{j+1,v}f - N_{j,v}f = (2^{\alpha+1}-1)A'(2^j)^{\alpha+1} \geq A'2^j \qquad \text{for } \alpha \geq 0.$$

The edge and regular part of the signal can be extracted out by the following condition,

$$N_{j+1,v}f - N_{j,v}f > \gamma,$$

where γ is a threshold that is appropriately selected.

Applications

Multiscale Edge Detection using WTMM (Canny's Edge Detector)

The edges of the different structures that appear in an image are often the most important features for pattern recognition. Edges points are often located where the image intensity has sharp transitions. Image textures do also have sharp intensity variations that are often not considered as edges. The discrimination of edges versus textures depends on the scale of analysis. This has motivated computer vision researchers to detect sharp image variations at different scales. So there is a large class of edge detectors look for points where the gradient of image intensity has a modulus which is locally maximum.

Canny's edge detector is a multiscale version of this approach [4]. Canny stated that a good edge detector should possess the following three properties: good

localization, high SNR and good continuity. The Canny operator works in a multi-stage process. First of all the image is smoothed by a convolution kernel $\theta(x)$ that is dilated. Since a wavelet with n vanishing moments can be written as the n^{th} order derivative of the function $\theta(x)$, the resulting wavelet transform is a multiscale differential operator. Derivatives of Guassian function are most often used to guarantee that all maxima lines propagate up to the finest scale. Then a simple 2-d first derivative operator is applied to the smoothed image to highlight regions of the image with high first spatial derivatives. Edges give rise to ridges in the gradient magnitude image. The algorithm then tracks along the top of these ridges and sets to zero all pixels that are not actually on the ridge top so as to give a thin line in the output, a process known as non-maximal suppression. The tracking process exhibits hysteresis controlled by two thresholds: *T1* and *T2*, with *T1* > *T2*. Tracking can only begin at a point on a ridge higher than *T1*. Tracking then continues in both directions out from that point until the height of the ridge falls below *T2*. This hysteresis helps to ensure that noisy edges are not broken up into multiple edge fragments.

Signal denoising using WTMM

WTMM carry the properties of sharp signal transitions and singularities. If one can reconstruct a signal from these maxima, it is then possible to modify the singularities of a signal by processing the WTMM. The strength of singularities can be modified

by changing the amplitude of the maxima and we can remove some singularities by suppressing the corresponding maxima. The interscale evolution property as shown in theorem 2.1 was first used in image denoising algorithms [10,39]. The noisy wavelet coefficients are detected by evaluating the Lipschitz exponents based on the maxima curve inside the COI in the scale-space plane.

Using the same wavelet kernel as the Canny edge detector, we perform the dyadic wavelet transform of a noisy signal. Then the Lipschitz exponents are estimated by tracing the maxima curves inside the COI in the scale-space plane. The maxima curves are removed or adjusted when they possess negative Lipschitz exponents. The denoised signal is then reconstructed from the processed WTMM by alternative projection algorithm such as the POCS technique [39].

Signal and image denoising using WTMS

In signal or image denoising applications, we use the interscale ratio and interscale difference of WTMS as the criteria to identify the noisy wavelet coefficients [32].

For 1-d signal, we perform the same dyadic wavelet decomposition. Then we compute $N_{j,v}f$ over the COI for all levels $1 \le j < J$. If $N_{j+1,v}f / N_{j,v}f > 2$ and $N_{j+1,v}f - N_{j,v}f \ge \gamma$, where γ is a threshold value, then we will select $W_j f(v)$ and the coefficients within the COI at the remaining scales. We repeat this process until all the coefficients corresponding to all the signal points have been processed. We then

reconstructed the denoised signal from the selected coefficients simply using the inverse wavelet transform.

For 2-d image, after the dyadic wavelet decomposition (in a non-separable way), we compute $N_{j,v}f$ over the *directional* COI for all levels $1 \le j < J$. Then select $W_j^1 f(v)$ and $W_j^2 f(v)$ if $N_{j+1,v}f / N_{j,v}f > 2$ and $N_{j+1,v}f - N_{j,v}f \ge \gamma$, where γ is a threshold value. Repeat until all the coefficients are processed. The denoised image can be reconstructed by simply applying the inverse wavelet transform.

2.3 Multiwavelet transform

Wavelet transforms with single scaling and wavelet functions have been studied extensively for more than a decade. Multiwavelet transform, with multiple scaling and wavelet functions operate in vector and matrix forms, have recently drawn much attention in the research community since the multiwavelet functions constructed by Geronimo *et al.* in 1994 have more desirable properties than any single wavelet transform. In this section, we briefly review the history and development, the important features and the realization of the multiwavelet transform.

2.3.1 History and development of the multiwavelet transform

Generalization of Scalar Wavelet

The study of multiwavelets was initiated by Goodman, Lee and Tang [22] in 1993. They extended Mallat's construction of orthonormal wavelets generated by a single function to multiwavelets generated by a finite set of functions. They found that there is a relationship between wavelets and the concept of wandering subspaces in operator theory. This provides a general setting to wavelets in Hilbert space where the translation and dilation operators are replaced by unitary operators. They proved the existence of orthonormal wavelet bases for V_m generated by a finite number of

vectors. They also proved the existence of bases of orthonormal wavelets for W_m and derive other results which are reminiscent of multiresolution approximation.

GHM Multiwavelets

Later in 1994, Goodman and Lee derived the necessary and sufficient conditions for the translates of $\phi_1,\cdots,\phi_r$, $\psi_1,\cdots,\psi_r$ to form a Riesz basis for V_1 [23]. The resulting decomposition and reconstruction sequences lead to the construction of dual bases for V_0 and its orthogonal complement W_0 in V_1. The general theory can be applied to the construction of spline wavelets with multiple knots (spline multiwavelets) [23,24]. Algorithms for the construction of these multiwavelets in some special cases are given in [23]. After that, Goodman constructed multiwavelets that have certain derivatives vanishing at the integers and satisfy certain interpolating conditions [24]. On the other hand, Geronimo, Hardin and Massopust proposed a method for constructing translation and dilation invariant function spaces using fractal functions defined by a certain class of iterated function systems in the same year [21]. These function spaces are generated by several scaling functions and their integer-translates. They also gave necessary and sufficient conditions for these function spaces which form a multiresolution analysis of $L^2(\Re)$. These multiwavelets are known as the GHM multiwavelets and the filter coefficient matrices are shown in appendix IV. In 1995, Strang and Strela carried out that the final step of the construction and produced

vector-valued wavelets by using paraunitary vector filter bank theory. In particular, they constructed vector-valued Meyer wavelets that are band-limited. They classified and constructed the vector-valued wavelets with sampling property. As an application, multiwavelets can be constructed from these vector-valued wavelets. They also showed that certain linear combinations of known scalar-valued wavelets can yield multiwavelets. In addition, they presented the discrete vector wavelet transform for discrete-time vector-valued (or blocked) signals, which can be thought of as a family of unitary vector transforms.

Prefiltering

It was found that in order to have a reasonable decomposition for discrete multiwavelet transforms, prefiltering is necessary [49,57]. A prefilter design method was introduced by Xia, Geronimo, Hardin and Suter in 1996 [57], where the idea is based on the computability of the multiwavelet transform coefficients from uniformly sampled signals. Moreover, an interpretation of the lowpass and highpass properties for vector filters was introduced. The criterion is, however, only good for the first step discrete multiwavelet transform decomposition. The prefilters designed with this method may be nonorthogonal, which might kill the gain of the energy compaction in the transform domain after the decoding is performed.

In [29], a different approach was proposed by Hardin and Roach for preserving

the orthogonality by using the approximation order criterion. In [35,92], Lebrun and Vetterli studied the *balanced* multiwavelets, where prefiltering for these kinds of multiwavelets is not necessary, but the other properties such as the short support and the smoothness, are not as good as the GHM multiwavelets.

In 1998, Xia introduced a new prefilter design technique which deals with all decomposition steps for the discrete multiwavelet transform [59]. The prefilter design technique is based on approximating a function with the lowpass property. The orthogonality of their translations is obtained from the linear combination of the multiscaling functions and their translations. The prefilter is not maximally decimated, which allows more freedom in designing. This prefilter was designed with the constraint that their combination with the multiscaling and multiwavelet filters will have lowpass and highpass property. Although it is orthogonal which is energy preserving, it is first order only. While Plonka's prefilter [43] is symmetric and second order. It is synthesized with perfect reconstruction filters having linear phase.

In 1998, Jiang introduced a procedure of designing orthogonal multiwavelets with good time-frequency resolution [34]. Formulas for computing the time-durations and the frequency-bandwidths of the scaling functions and multiwavelets are also obtained. Parameter expressions for the matrix coefficients of the vector filter banks that generate symmetric/antisymmetric scaling functions and multiwavelets supported in

$[0, N]$ were presented for $N = 2, \cdots, 6$. Moreover, orthogonal multiwavelets with optimum time-frequency resolution were constructed with some optimal multifilter banks provided. All the popular prefilter coefficient matrices are shown in the appendix IV.

2.3.2 Important features of the multiwavelet transform

One may interest at the features of the multiwavelets that the single wavelet cannot possess simultaneously. It was verified by Strang and Strela in 1995 [48] that for any single function ϕ apart from the Haar's piecewise constants, it is impossible that ϕ is symmetric (linear phase), short support (two intervals or less), and their translation form an orthogonal family. However, those properties can be readily achieved by introducing 2×2 matrix coefficients. They also demonstrated the nice properties of the multiwavelets, for example, the scaling functions derived by Geronimo, Hardin and Massopust in 1994 [21]. The properties are shown as the following.

1. *Symmetry:* Φ_1 *and* Φ_2 *are even after a shift of the origin, that is, they are linear phase filters.*
2. *Short support:* Φ_i *vanishes outside the interval* $[0, i]$. *GHM scaling functions have supports* $[0,1]$ *and* $[0,2]$ *and their supports are slightly shorter than CL multiwavelet filters.*

3. *Orthogonality: The translates* $\Phi_1(t-k)$ *and* $\Phi_2(t-k)$ *are all mutually orthogonal.*

4. *Regularity (approximation order and smoothness): Classical notions of regularity are the approximation power and smoothness. GHM multifilters have second-order approximation, (linear combination of the scaling and wavelet functions can reproduce constant and linear functions). CL multifilters have third-order approximation. GHM multifilters are smoother than CL multifilters.*

5. *Vanishing moment (order of multiscale differential operator): Multiscale differential operator is used to produce multiscale modulus maxima which locate discontinuities of a signal. One vanishing moment: the modulus maxima will be the maxima of the first order derivative of the signal smoothed by the kernel function; two vanishing moments: the modulus maxima will correspond to high curvatures. The more vanishing moments the filters have, the smaller transform coefficients will be produced at fine scales. GHM multiwavelets have enough vanishing moments.*

Multiresolution Approximation

Let us define a multiresolution analysis of $L^2(\Re)$ generated by several scaling functions, with an increasing sequence of function subspaces $\{V_j\}_{j\in Z}$ in $L^2(\Re)$,

$$\{0\} \subset \cdots \subset V_{-1} \subset V_0 \subset V_1 \subset \cdots \subset L^2(\Re). \tag{2.35}$$

Subspaces V_j are generated by a set of scaling functions, namely, multiscaling functions, such that V_j is the closure of the linear span of $\{\phi_{l,k}^{(j)}\}_{1\le l\le r,k\in Z}$ in $L^2(\Re)$, where

$$\phi_{l,k}^{(j)}(t)=2^{j/2}\phi_l(2^j t-k), \qquad \forall t\in\Re. \tag{2.36}$$

Then we have a sequence of multiresolution subspaces $\{V_j\}_{j\in Z}$ generated by a set of multiscaling functions, where the resolution gets finer and finer as j increases.

Let us define inter-spaces $W_j\subset L^2(\Re)$ such that $V_{j+1}=V_j\oplus W_j$, $\forall j\in Z$. W_j is the complement of V_j in V_{j+1} and thus W_j and W_l with $j\neq l$ may not be orthogonal to each other. Subspaces W_j can be generated by r base functions that are the multiwavelets $\psi_1,\psi_2,\cdots,\psi_r$. The subspaces W_j is the closure of the linear span of $\{\psi_{l,k}^{(j)}\}_{1\le l\le r,k\in Z}$, where

$$\psi_{l,k}^{(j)}(t)=2^{j/2}\psi_l(2^j t-k), \qquad \forall t\in\Re. \tag{2.37}$$

2.3.3 Realization of the multiwavelet transform

The pyramid algorithm for computing the single wavelet transform coefficients can be implemented by using tree-structured multirate filter banks. But for the multiwavelets, in 1996, Xia, Geronimo, Hardin and Suter proposed a general algorithm to compute the multiwavelet transform coefficients by adding proper pre-multirate filter bank (or prefilter) [57] before the vector filter banks that generate multiwavelets [58]. This

preprocessing is usually performed before the multiwavelet decomposition. It can be thought as the vectorization of an input signal to produce a certain discrete-time vector-valued signals for the discrete vector-valued wavelet transform. After the multiwavelet reconstruction, postprocessing (or postfiltering) is applied to give a one-stream recovered signal. Preprocessing does not exist in single wavelet transform. Prefiltering often produce correlated coefficients which contain information of the regularity of the input signal. Multifiltering captures this information and further characterizes it with its vector-valued transform.

Some numerical experiments were also done by them to illustrate the energy compaction of the combined discrete multiwavelet transform could be higher than the conventional discrete single wavelet transform. In this part, we briefly review the implementation of the discrete multiwavelet transform and the perfect reconstruction conditions. A generalized prefilter design was proposed by Xia in [59]. It is orthogonal but is 1^{st} order only. Plonka's prefilter is 2^{nd} order and symmetric [43], while Yang *et. al* 's prefilter is 2^{nd} order and orthogonal [60]. We review the realization of multiwavelet transform combined with the pre- and postfiltering in [57].

Consider N general orthogonal wavelets with compact support, that is N compactly supported scaling functions $\phi_l(t)$, $l = 1,2,\cdots,N$, and N mother wavelet functions $\psi_l(t)$, $l = 1,2,\cdots,N$, where $\phi_l(t-k)$, $k \in Z$, $l = 1,2,\cdots,N$ which are

mutually orthogonal. $2^{j/2}\psi_l(2^j t-k)$, $j,k\in Z$, $l=1,2,\cdots,N$ form an orthonormal basis for $L^2(\Re)$. Let $\mathbf{H}(\omega)$ and $\mathbf{G}(\omega)$ be their corresponding $N\times N$ matrix quadrature mirror filters with impulse response H_k, and G_k, $k\in Z$, respectively. Let $\Phi(t)\equiv(\phi_1(t),\cdots,\phi_N(t))^T$ and $\Psi(t)\equiv(\psi_1(t),\cdots,\psi_N(t))^T$. Then $\Phi(t)$ and $\Psi(t)$ can be expressed by the following matrix dilation equations,

$$\Phi(t)\equiv 2\sum_{\forall k} H_k\Phi(2t-k), \tag{2.38}$$

and

$$\Psi(t)\equiv 2\sum_{\forall k} G_k\Phi(2t-k). \tag{2.39}$$

For each fixed $j\in Z$, let V_j be the closure of the linear span of $2^{j/2}\phi_l(2^j t-k)$, $l=1,2,\cdots,N$, $k\in Z$. Then, the spaces V_j, $j\in Z$ form an orthogonal multiresolution analysis for $L^2(\Re)$. Although they only focus on two-band multiwavelets, the theory developed can be easily generalized to M-band wavelets where there are N scaling functions and $(M-1)N$ mother wavelet functions.

Let $f\in V_0$, then

$$f(t)=\sum_{l=1}^{N}\sum_{k\in Z} c_{l,0,k}\phi_l(t-k), \tag{2.40}$$

$$\Rightarrow f(t)=\sum_{l=1}^{N}\sum_{k\in Z} c_{l,J_0,k}2^{J_0/2}\phi_l(2^{J_0}t-k)+\sum_{l=1}^{N}\sum_{J_0\le j\le 0}\sum_{k\in Z} d_{l,j,k}2^{j/2}\psi_l(2^j t-k), \tag{2.41}$$

where $J_0<0$ and $c_{l,j,k}$, $d_{l,j,k}$ are defined as

$$c_{l,j,k}=\int f(t)2^{j/2}\phi_l(2^j t-k)dt,$$

and

$$d_{l,j,k}=\int f(t)2^{j/2}\psi_l(2^j t-k)dt.$$

Let $\mathbf{c}_{j,k}\equiv(c_{1,j,k},\cdots,c_{N,j,k})^T$ and $\mathbf{d}_{j,k}\equiv(d_{1,j,k},\cdots,d_{N,j,k})^T$. Then, the decomposition and

reconstruction formulas are

$$\mathbf{c}_{j-1,k} \equiv \sqrt{2}\sum_{\forall n} H_n \mathbf{c}_{j,2k+n}, \tag{2.42}$$

$$\mathbf{d}_{j-1,k} \equiv \sqrt{2}\sum_{\forall n} G_n \mathbf{d}_{j,2k+n}, \tag{2.43}$$

and

$$\mathbf{c}_{j,n} \equiv \sqrt{2}\sum_{\forall k}\left(H_k \mathbf{c}_{j-1,2k+n} + G_k \mathbf{d}_{j-1,2k+n}\right). \tag{2.44}$$

Therefore, to determine thce wavelet coefficients $\mathbf{c}_{J_0,k}$ and $\mathbf{d}_{j,k}$ for $J_0 \le j < 0$, $k \in Z$ from the samples of $f(t)$, it is only necessary to determine the coefficients $\mathbf{c}_{0,k}$ for $k \in Z$ from the samples of $f(t)$.

Suppose the samples $f\left(\frac{n}{M}\right)$ of $f(t)$ with sampling rate $\frac{1}{M}$ are known. Let

$$x[n] \equiv f\left(\frac{n}{M}\right),\ n \in Z, \qquad X(\omega) = \sum_{\forall n} x[n] e^{-i\omega n},$$

$$X_m(\omega) = \sum_{\forall n} x[Mn+m] e^{-i\omega n}, \tag{2.45}$$

where $X_m(\omega)$ is the m th polyphase component of $X(\omega)$ for $m = 0,1,\cdots,M-1$.

Let $$P_{m,l}(\omega) = \sum_{\forall n} \phi_l\left(\frac{m}{M}+n\right) e^{-i\omega n}, \quad l = 1,2,\cdots,N,\ m = 0,1,\cdots,M-1, \tag{2.46}$$

and $$C_{l,j}(\omega) = \sum_{\forall k} c_{l,j,k} e^{-i\omega k}, \qquad l = 1,2,\cdots,N,\ j \in Z. \tag{2.47}$$

Let $\mathbf{P}(\omega)$ be the following $M \times N$ matrix function

$$\mathbf{P}(\omega) = \begin{pmatrix} P_{0,1}(\omega) & P_{0,2}(\omega) & \cdots & P_{0,N}(\omega) \\ P_{1,1}(\omega) & P_{1,2}(\omega) & \cdots & P_{1,N}(\omega) \\ \vdots & \vdots & \ddots & \vdots \\ P_{M-1,1}(\omega) & P_{M-1,2}(\omega) & \cdots & P_{M-1,N}(\omega) \end{pmatrix}. \tag{2.48}$$

The Fourier transform of eqn.2.40 with $t = \frac{m}{M} + n$, $m = 0,1,\cdots,M-1$ yields

$$\left(X_0(\omega) \ \cdots \ X_{M-1}(\omega)\right)^T = \mathbf{P}(\omega)\left(C_{1,0}(\omega) \ \cdots \ C_{N,0}(\omega)\right)^T. \tag{2.49}$$

Since all $\phi_l(t)$, $l = 1,2,\cdots,N$ are assumed to have compact supports and all the

entries of the matrix function $\mathbf{P}(\omega)$ are polynomials of $e^{-i\omega}$, by eqn.2.49, the following proposition gives the solvability of the transform coefficients from samples.

Proposition 2.2

The wavelet transform coefficients $\mathbf{c}_{J_0,k}$ *and* $\mathbf{d}_{j,k}$ *for* $J_0 \le j < 0$, $k \in Z$ *can be exactly computed from* $f\left(\frac{n}{M}\right)$, $n \in Z$, *if and only if* $M \ge N$ *and the rank of the matrix* $\mathbf{P}(\omega)$ *in eqn.2.54 is always* N *for all* $\omega \in [0,2\pi)$.

Since efficient sampling is used, the sampling rate should be as small as possible. So based on proposition 2.2, we can assume $M = N$ and have these corollaries.

Corollary 2.1

The wavelet transform coefficients $\mathbf{c}_{J_0,k}$ *and* $\mathbf{d}_{j,k}$ *for* $J_0 \le j < 0$, $k \in Z$ *can be exactly computed from* $f\left(\frac{n}{N}\right)$, $n \in Z$, *if and only if the determinant function of the matrix function* $\mathbf{P}(\omega)$ *does not have any zeros for* $\omega \in [0,2\pi)$, *that is, the inverse of* $\mathbf{P}(\omega)$ *exists.*

Corollary 2.2

The inverse of $\mathbf{P}(\omega)$ *is FIR if and only if the determinant of* $\mathbf{P}(\omega)$ *is* $ce^{im_0\omega}$ *for a certain nonzero constant* c *and a certain integer* m_0.

Let $\mathbf{Q}(\omega)$ be the inverse of $\mathbf{P}(\omega)$, that is, $\mathbf{P}(\omega)\mathbf{Q}(\omega) = I_N$. The decomposition and reconstruction of $\mathbf{c}_{j,k}$ and $\mathbf{d}_{j,k}$ from $f\left(\frac{n}{N}\right)$ is shown in figure 2.5.

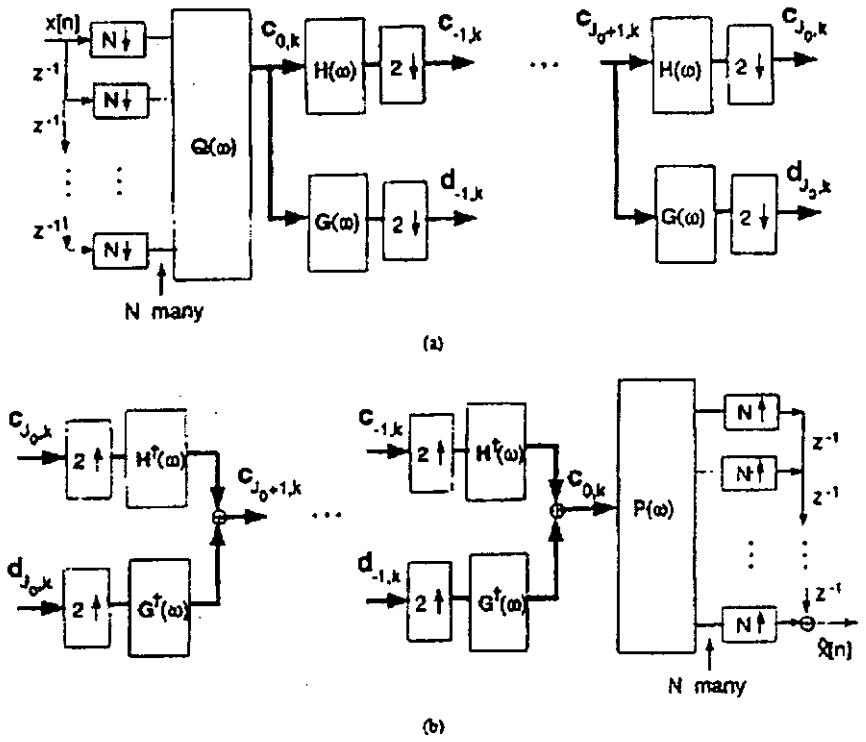


Fig 2.5 Discrete multiwavelet filter banks with Xia's prefilter [91]. (a) Decomposition (b) Reconstruction.

Proposition 2.3

The system in figure 2.5 is perfect reconstruction, that is, $\hat{x}[n] = x[n]$, *if and only if the matrix quadrature mirror filters* $\mathbf{H}(\omega)$ *and* $\mathbf{G}(\omega)$ *satisfy*

$$\mathbf{H}(\omega)\mathbf{H}^*(\omega) + \mathbf{H}(\omega+\pi)\mathbf{H}^*(\omega+\pi) = I_N, \tag{2.50}$$

$$\mathbf{G}(\omega)\mathbf{G}^*(\omega) + \mathbf{G}(\omega+\pi)\mathbf{G}^*(\omega+\pi) = I_N, \tag{2.51}$$

$$\mathbf{H}(\omega)\mathbf{G}^*(\omega) + \mathbf{H}(\omega+\pi)\mathbf{G}^*(\omega+\pi) = 0_N, \tag{2.52}$$

and

$$\mathbf{P}(\omega)\mathbf{Q}(\omega) = I_N. \tag{2.53}$$

When a discrete-time signal $x[n]$ is considered, proposition 2.3 also suggests that one may use other pre- or post-filters $\mathbf{Q}(\omega)$ and $\mathbf{P}(\omega)$ rather than the one in

eqn.2.49. So for fixed $\mathbf{H}(\omega)$ and $\mathbf{G}(\omega)$, there are many algorithms in terms of different $\mathbf{Q}(\omega)$ and $\mathbf{P}(\omega)$. Which one is good? The combined filters $\mathbf{H}(\omega)\mathbf{Q}(\omega)$ and $\mathbf{G}(\omega)\mathbf{Q}(\omega)$ that have lowpass and bandpass properties respectively are preferred. Since a $N\times N$ matrix filter $\mathbf{F}(\omega)$ is also a polyphase matrix of N filters $F_1(\omega),\cdots,F_N(\omega)$, one way to interpret the lowpass and bandpass properties of a matrix filter $\mathbf{F}(\omega)$ is to use the lowpass and bandpass properties of its associated N filters $F_l(\omega)$, where $l=1,2,\cdots,N$. Let $\mathbf{H}(\omega)\mathbf{Q}(\omega)$ be the polyphase matrix of $\hat{H}_l(\omega)$, $l=1,2,\cdots,N$ and $\mathbf{G}(\omega)\mathbf{Q}(\omega)$ be the polyphase matrix of $\hat{G}_l(\omega)$, $l=1,2,\cdots,N$.

Assume $$\hat{H}_l(\pi)=0,\ l=1,2,\cdots,N \tag{2.54}$$

and $$\hat{G}_l(\pi)=0,\ l=1,2,\cdots,N. \tag{2.55}$$

If the prefilter $\mathbf{Q}(\omega)$ that have the inverse (or the postfilter) $\mathbf{P}(\omega)$, has determinant $\det(\mathbf{Q}(0))=\pm 1$, and satisfies eqn.2.54 and eqn.2.55, then $\mathbf{Q}(\omega)$ are called good prefilters with respect to $\mathbf{H}(\omega)$ and $\mathbf{G}(\omega)$.

2.4 Summary

In this chapter, we reviewed the history and development of the wavelet and multiwavelet transform. We highlighted their important features, such as time-frequency localization and multiresolution approximation, which are crucial to the singularity detection and the algorithms proposed in the next two chapters. We also reviewed the realization of them by discrete filter bank and multifilter bank respectively with perfect reconstruction. For the wavelet transform, we reviewed the background of the singularity detection. For the multiwavelet transform, we will present it in chapter 4 because it was not developed before.

Chapter 3

Regularity Scalable Wavelet Image Coding

3.1 Introduction

In this chapter, we suggest making use of the estimation of Lipschitz regularity to achieve an efficient regularity based scalable image coding algorithm. In section 2.2.4, Lipschitz regularity was proven to be characterized by the wavelet transform for image analysis applications such as pattern recognition and computer vision. In current wavelet image coders, wavelet coefficients are often computed by fast filter bank algorithm, where separable discrete convolutions are usually performed, followed by subsamplings. This is the separable wavelet transform with decimation we refer to here, and we need to check whether it is plausible to estimate the Lipschitz regularity from the evolution of the separable wavelet transform across scales at all orientations. So we proved the Lipschitz regularity condition for the separable wavelet transform [30]. Then based on this, we proposed a scalable wavelet image coding algorithm [30,86] which utilizes tree structured data organization. The comparison with the existing feature-based scalable wavelet image coding algorithms showed that the decoded images are higher in PSNR and have better visual quality at the same bit

rates. Significant improvement of the edges and textures components of a decoded image can be shown at lower bit rates. When we combine this scalable coding algorithm with resolution scalability, the bit rates of a decoded image can be greatly reduced at the same quantization level. Although the PSNR is also decreased, the decrement is very small and the change in visual quality is negligible.

The organization of this chapter is the following. We first review the existing scalable wavelet image coding algorithms, which are based on the image features. Then we prove the Lipschitz regularity condition for the separable wavelet transform. Since wavelet transform magnitude sum developed before cannot be directly applied in this case, we determine the decimated COI for the magnitude sum of the separable wavelet transform. We show how it is used to detect singularities or estimate the Lipschitz regularity. Next we propose a simple, efficient and adaptive regularity scalable image coding algorithm based on this. Finally, the results and the improvements over the several existing algorithms are shown, followed by a short summary.

3.2 The present works

With the decreasing cost and increasing performance of digital multimedia capture devices, computing power and storage capabilities, more and more visual information will be available on-line in large image and video repositories [101]. For rapid transmission or fast image browsing from the database, if there is insufficient bandwidth in the network, packets will be dropped indiscriminately during congestion. Even with the advanced resource management QoS, there are still many irresponsible error condition and non-scalable state information on a per-flow granularity. Therefore, one should provide a fast coarse approximation of the image, and then progressively enhance it as more bits are transmitted and received. This fascinates numerous signal processing researchers to design and implement scalable image and video coding algorithms for these multimedia communication applications, so that at the same time, the target bit-rate need not to be known at the time of compression.

3.2.1 Resolution and SNR scalable wavelet image coding

Current well-known wavelet image compression algorithms such as EZW algorithm [47] in MPEG-4 [104], SPIHT algorithm [45] and EBCOT algorithm [50] in JPEG2000 [105] provide scalabilities in resolution and SNR or accuracy. Resolution

scalability (figure 3.1) allows different resolutions or sizes of a decoded image to be displayed, mostly affects the bit rates. SNR or accuracy scalability (figure 3.2) allows images with different numbers of significant bit-planes decoded, while the total number of encoded bit-planes, which represents the precision of the decoded wavelet coefficients allowed, is regulated by the quantizer of the image coder. These scalabilities only allow a coarse to fine approximation of images at various bit rates. However, these quality levels are produced without any correlation with the visual perception, because they are not scalable on the features of an image. Naturally, edges, boundaries, textures and surfaces are considered to be the visually important features of an image. The visual quality of an image will be improved if the information of these features is increased. So Scalability on these features can provide a better visual appearance of an image at very low bit rates, as the display of the shape or the visually important features in a decoded image is allowed first at very low bit rate. It can also emphasize a particular feature of an image at a certain bit rate.

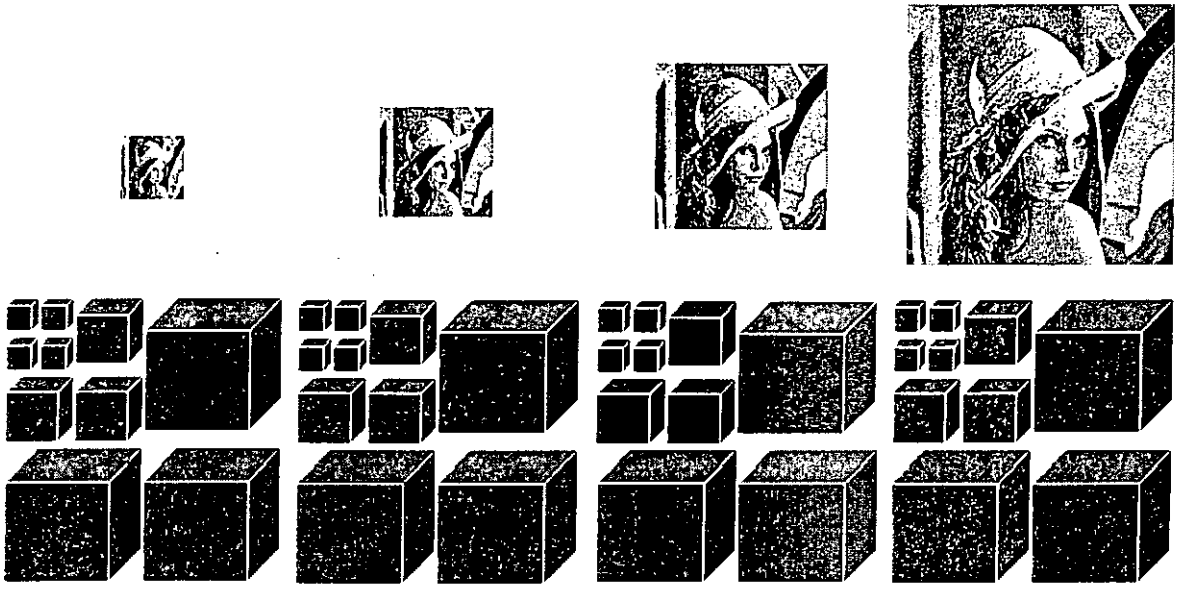


Figure 3.1 Illustration of Resolution scalability.

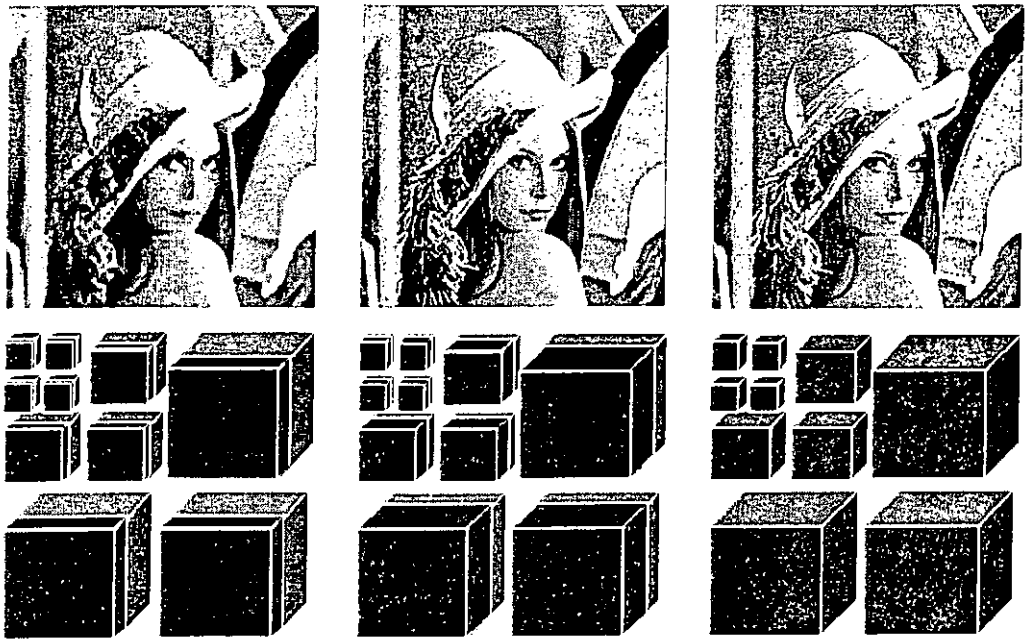


Figure 3.2 Illustration of SNR scalability.

3.2.2 Edge-enhanced wavelet image coding

For image browsing and retrieval over a bandwidth varying network, it is desirable to display or recognize the shape of an image at very low bit rates, and then progressively enhance it as more bits are received. So various progressive or scalable wavelet image coding algorithms based on edges and surfaces of an image were proposed [91,94,97]. Before going to the scalable coding algorithms, let us investigate the other approaches, which can achieve the above objectives. In [97], an edge-enhanced image coding algorithm which combines the edge detection algorithm SUSAN with the SPIHT algorithm [45] was proposed (figure 3.3). However, the combination is not simple for the edge-enhanced reconstruction, so it is complicated to provide scalability on the edges based on this algorithm, though it improves the recognizability of decoded images at very low bit rates (figure 3.4), and the edge detection algorithm itself is insensitive to noise and has a good edge localization. Moreover, the multiring chain code used is particularly effective for long and smooth curves only. So the edge coding is too costly for most images.

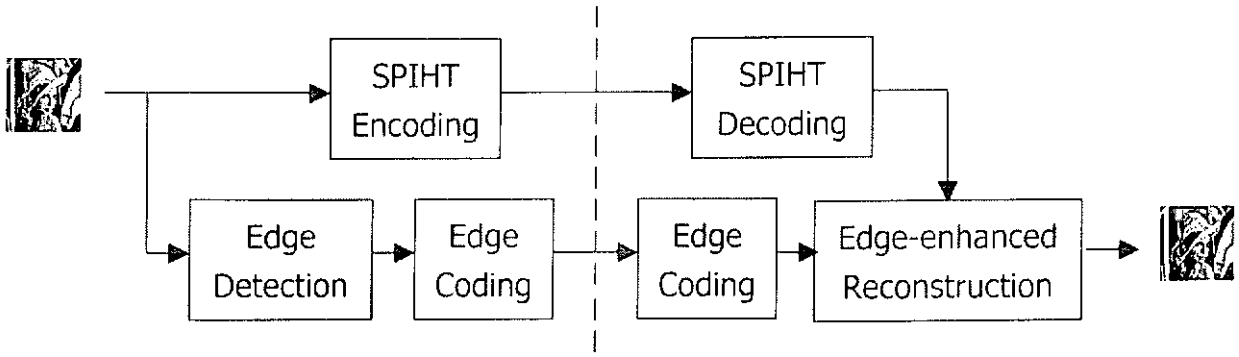
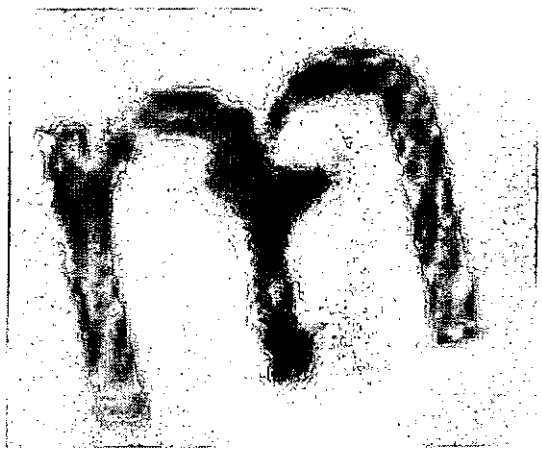
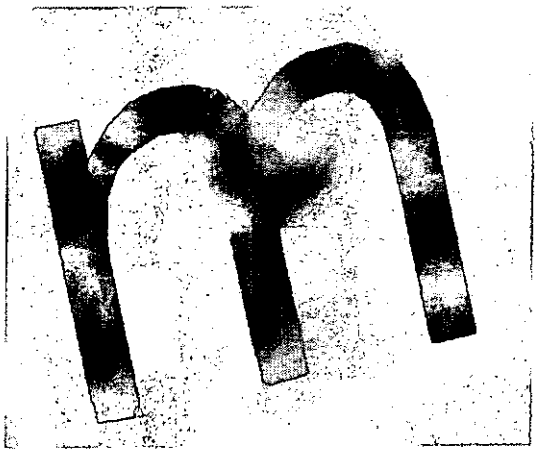


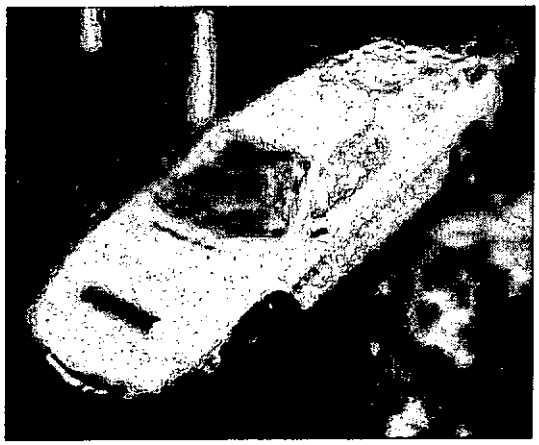
Figure 3.3 Edge-enhanced image coding algorithm in [97].



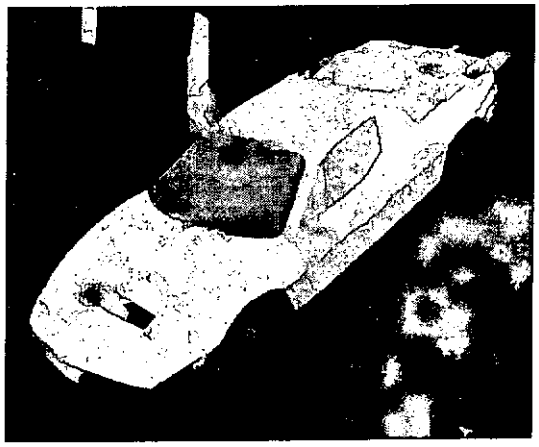
0.0099bpp, SPIHT coded



0.0099bpp, edge-enhanced
(0.0049bpp for edge coding)



0.0298bpp, SPIHT coded



0.0298bpp, edge-enhanced
(0.0128bpp for edge coding)

Figure 3.4 Edge-enhanced images using algorithm in [97].

3.2.3 Edge-oriented scalable wavelet image coding

Traditional scalable coding techniques encode images starting from the lowest resolution subband. One may encounter a situation that, at very low bit rate, the low frequency information together with the blurred and blotchy artifacts will dominate the decoded image. In hierarchical octave-band wavelet image coding, wavelet coefficients at different scales and orientations can produce different effects to these features. To perform scalability on these features, we need to quantify them with critical determination of how to select the appropriate coefficients from their wavelet decomposition, how to combine them and in what sequence to encode and transmit them. In [91], another progressive coding scheme was proposed (figure 3.5). The wavelet transform is used to produce a multiple resolution framework so that high resolution and visually important components of images such as edges is sent first, and the skeleton shape of a decoded image is outlined at low bit rates (figure 3.6). This can be achieved by encoding and transmitting subbands starting from the highest resolution one. Although efficiency can be improved without any significant effects to a decoded image by discarding the first highest resolution subband, numerous bits are still needed to encode the higher resolution subbands first for the simplest edges. Moreover, the impression of the decoded images is generally not good because the gray level is not displayed almost until all the subbands are received and decoded.

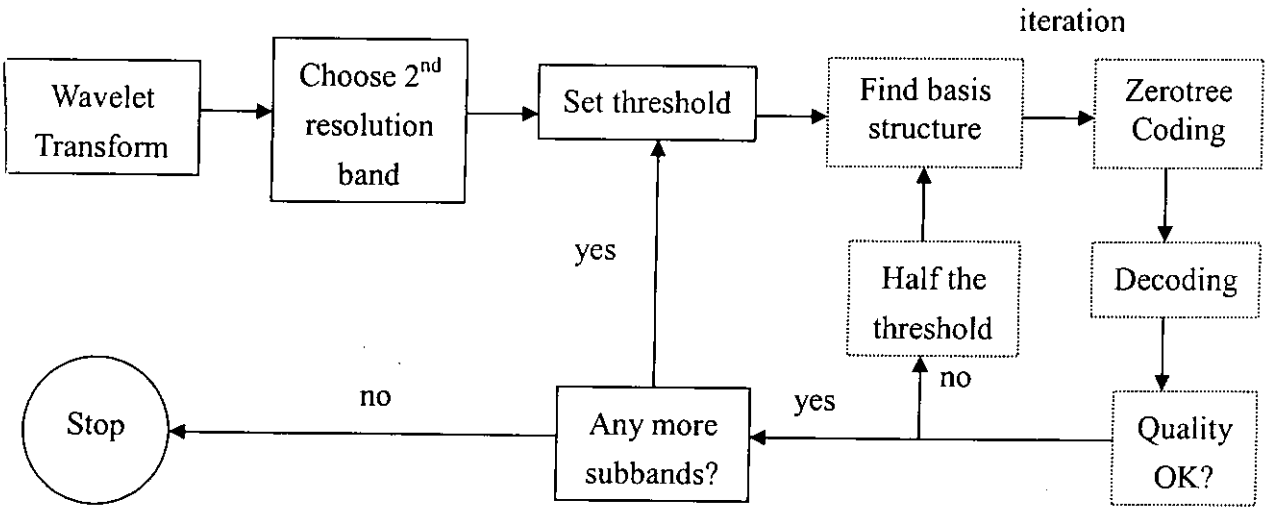


Figure 3.5 Edge-oriented progressive image coding algorithm in [91].

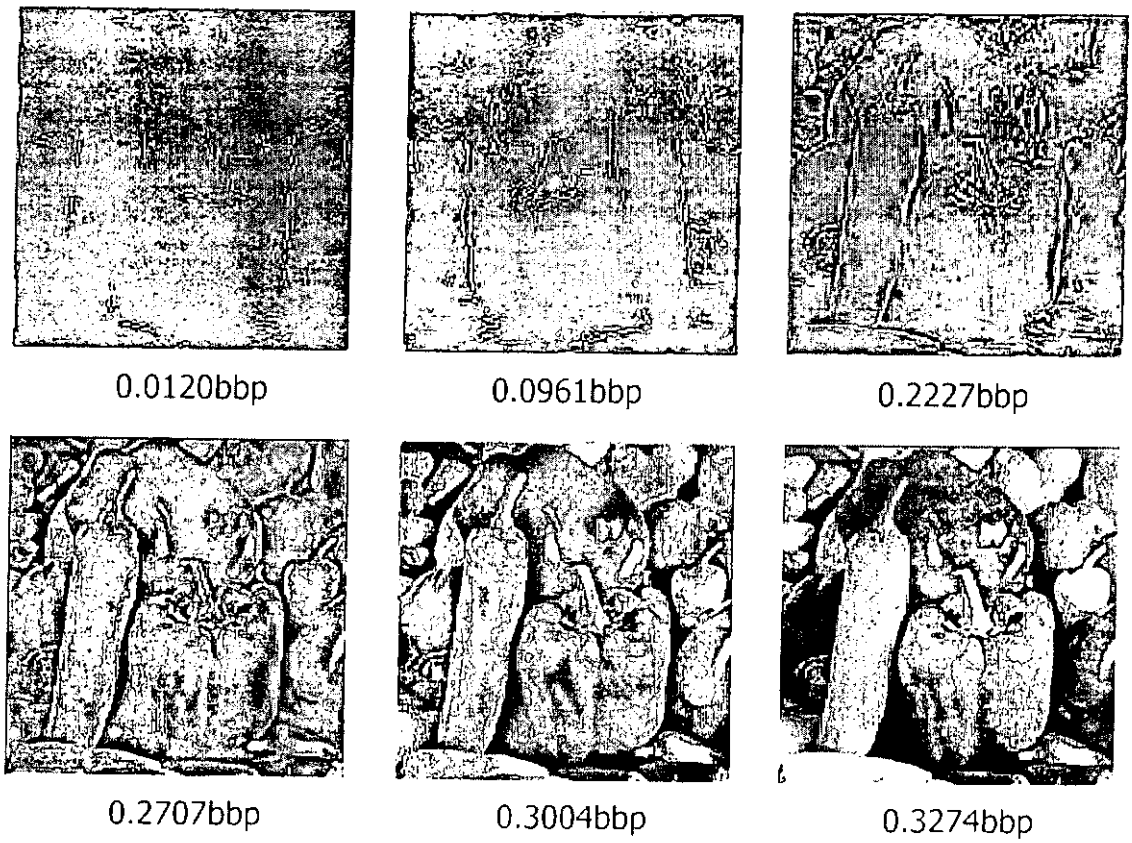


Figure 3.6 Results of the edge-oriented progressive image coding algorithm in [91].

3.2.4 Surface-oriented scalable wavelet image coding

According to the properties of HVS, surface orientation is important for recognition and extraction of visual objects. Therefore, in [94], a scalable coding algorithm based on surface orientation was proposed. It was suggested that the gray levels should be maintained at all time for a 'good impression' of the decoded images. The surface orientation from the shadings of images is used as the criteria to decide the ordering of the subbands and their bit-planes for encoding and transmission (figure 3.7). The surface orientation is measured by the stereographic projection of the reflectance map obtained from the image brightness. With this algorithm, coarse shapes of objects are first decoded, with shading information added progressively. This avoids using many bits for edge coding like the above two algorithms. The sequence of encoding the subbands at different bit-planes is shown in figure 3.8. However, since the complicated mathematical relation between the wavelet coefficients at different scales and orientations and the surface orientation is still not concrete, only a trial method can be utilized and thereafter the surface orientation is evaluated. Moreover, due to the assumption of point light source and lambertian object surfaces, the scope of images suitable for this algorithm is limited. Furthermore, when the bit rate is higher, the recovered images become more similar to the original one (figure 3.9), and it becomes harder to determine the transmission sequence as the surface orientations of

the images are similar. With these unsolvable problems, it still cannot be a desirable feature-based scalable wavelet image coding algorithm.

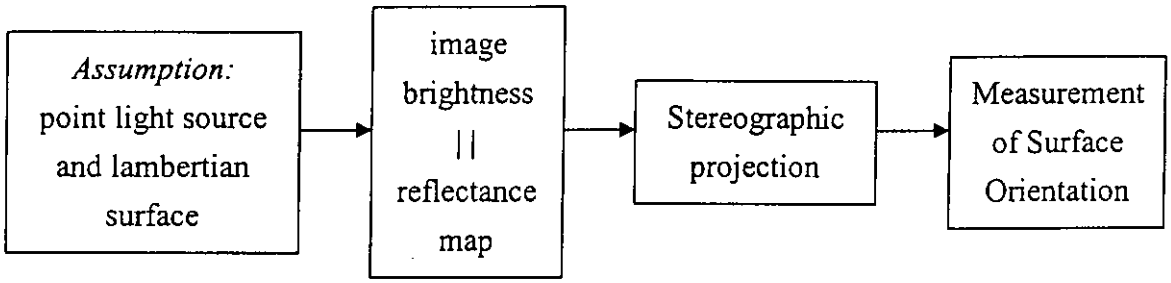
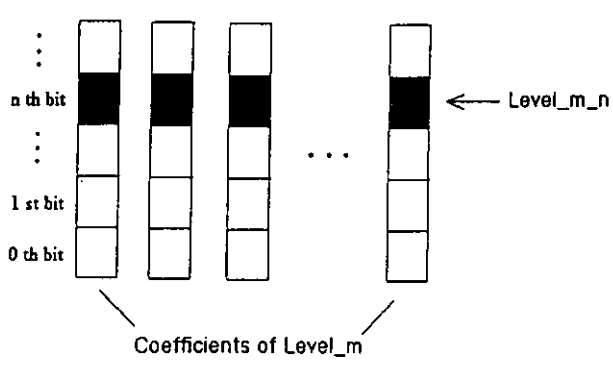


Figure 3.7 Surface-oriented scalable image coding algorithm in [94].



Level_6_8, Level_6_7, Level_5_6, Level_5_5, Level_6_6, Level_6_5, Level_6_4, Level_6_3, Level_6_2, Level_6_1, Level_5_4, Level_5_3, Level_5_2, Level_5_1, Level_4_6, Level_4_5, Level_3_6, Level_4_4, Level_4_3, Level_3_5, Level_3_6, Level_4_2, Level_4_1, Level_2_6, Level_2_5, Level_3_3, Level_2_4, Level_3_2, Level_3_1, ...

Figure 3.8 Sequence of encoding and transmitting the subbands at different bit-planes determined by the scalable image coding algorithm in [94].

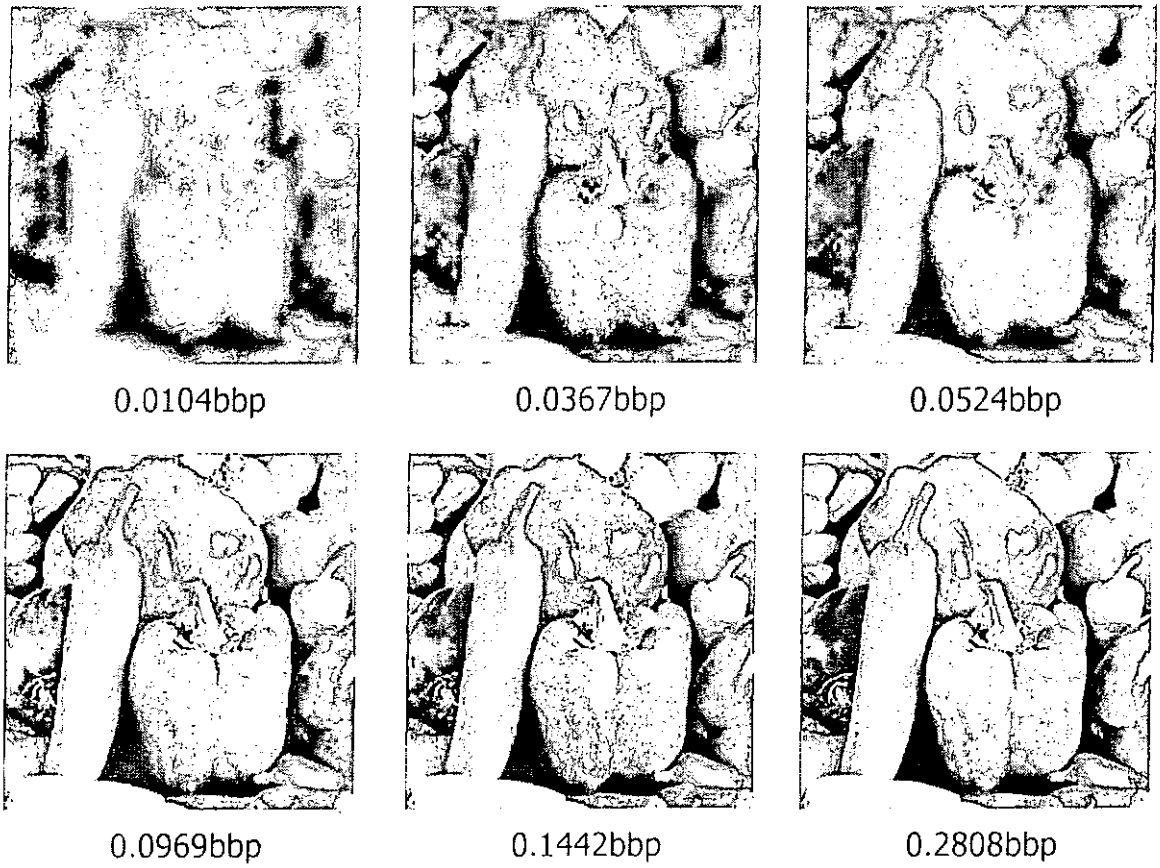


Figure 3.9 Results of the surface-oriented progressive image coding algorithm in [94].

3.3 The proposed algorithm

There are many nice features to embed the scalable coding algorithm into the current wavelet based image codec, where separable wavelet transform with decimation is performed. So we first introduce the separable wavelet transform and the Lipschitz condition [30]. Then we determine the decimated cone of influence (COI) for the magnitude sum and present the proposed scalable coding algorithm [30,86].

Definition 3.1 ***Separable Wavelet Transform***

Let $1 \le l \le 3$ *be the indices of the orientations and denote again* $x = (x_1, x_2)$ *and* $u = (u_1, u_2)$. *Let* ϕ *be a scaling function and* ψ *be the corresponding wavelet generating a wavelet orthonormal basis of* $L^2(\Re)$. *A separable wavelet orthonormal basis* $\{\psi^l\}_{1 \le l \le 3}$ *is then constructed with tensor products of the scaling function* ϕ *and the wavelet function* ψ. *That is,* $\psi^1(x) = \phi(x_1)\psi(x_2)$, $\psi^2(x) = \psi(x_1)\phi(x_2)$ *and* $\psi^3(x) = \psi(x_1)\psi(x_2)$. *Let* $\psi^l_{j,u}(x)$ *be the set of translations and dilations of the mother wavelet* $\{\psi^l\}_{1 \le l \le 3}$, $\psi^l_{j,u}(x) = 2^{-j}\psi^l(2^{-j}x_1 - u_1, 2^{-j}x_2 - u_2)$, *so that* $\{\psi^l_{j,u}\}_{1 \le l \le 3}$ *is an orthonormal basis of* $L^2(\Re^2)$ *(figure 3.10). Let* $\phi^2_{j,u}(x) = 2^{-J}\phi^2(2^{-J}x_1 - u_1, 2^{-J}x_2 - u_2)$, *where* $\phi^2(x) = \phi(x_1)\phi(x_2)$ *and* J *is the total number of scales. Then the separable wavelet transform of a signal* $f \in L^2(\Re^2)$ *at* u *is*

$$\left\{S_J f(u) = \langle f, \phi^2_{J,u} \rangle, \left(W^l_j f(u) = \langle f, \psi^l_{j,u} \rangle\right)_{1 \le j \le J, 1 \le l \le 3}\right\}. \qquad (3.1)$$

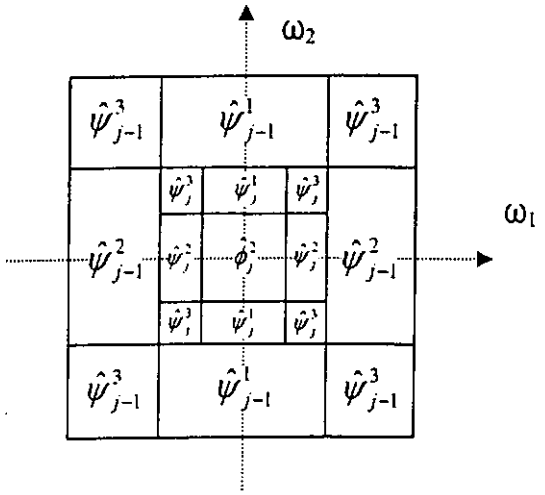


Figure 3.10 Dyadic rectangles indicate the frequency regions for which the energy of the $\psi^l_{j,u}(x)$ is mostly concentrated.

Theorem 3.1

If f is uniformly Lipschitz $\alpha \le n$ in the neighborhood of v, $U(v)$, there exists a finite constant such that

$$\forall (j,u) \in S^{(j,v)}, \quad \left| W_j^l f(u) \right| \le A\left(2^j\right)^{\alpha}. \tag{3.2}$$

Conversely, suppose that f is bounded and $\alpha < n$ is not an integer. If there exists a finite generic $A' > 0$ and $\alpha' < \alpha$ such that

$$\forall (j,u) \in S^{(j,v)}, \quad \left| W_j^l f(u) \right| \le A\left(2^j\right)^{\alpha'}, \tag{3.3}$$

then f is Lipschitz α in $U_\delta(v)$, where the region

$$U_\delta(v) = \left\{ u : |u_1 - v_1| \le \varepsilon + \delta, |u_2 - v_2| \le \varepsilon + \delta \right\}, \quad for \ \delta > 0,$$

belongs to the support of $\psi^l_{j,u}(x)$.

The proof is shown in Appendix I.

By theorem 3.1, it is plausible to characterize Lipschitz regularity from the evolution of the separable wavelet transform across scales. Experimental observations indicate that strong edges achieve equality in eqn.3.2 [5]. This is the extreme case of the condition in eqn.3.2 which we can distinguish different levels of Lipschitz regularity by computing the interscale ratios of the coefficient magnitudes. Therefore, one may suggest that different extents of regularity can be obtained from the selection of the decimated wavelet coefficients by the interscale ratios of their magnitudes. Wavelet zero-trees take advantage of the decay of wavelet coefficients by relating these coefficients across scales with quad-trees [36]. For each orientation, $l=1,2,3$, we create quad-trees by relating each coefficient $W_j^l f[u]$ to the following four children at the next finer scale 2^{j-1}, that is $W_{j-1}^l f[2u_1, 2u_2]$, $W_{j-1}^l f[2u_1-1, 2u_2]$, $W_{j-1}^l f[2u_1, 2u_2-1]$ and $W_{j-1}^l f[2u_1-1, 2u_2-1]$. At the largest scale 2^J, the children of $S_J f[u]$ are defined as the three wavelet coefficients at the same scale and location $W_J^1 f[u]$, $W_J^2 f[u]$ and $W_J^3 f[u]$. The construction of the quad-trees is illustrated in figure 3.11. The value of a wavelet coefficient and its four children depends on the variation of the image gray level in that spatial area. So we compute the interscale ratios as the following,

$$R_{j,u}^l f = \frac{\left|W_j^l f[u]\right|}{\sum_{u_1'=0}^{1}\sum_{u_2'=0}^{1}\left|W_{j-1}^l f[2u_1-u_1', 2u_2-u_2']\right|}, \qquad (3.4.1)$$

and at the coarsest scale $j = J$,

$$R^l_{J,u} f = \frac{|S_J f[u]|}{|W^l_J f[u]|}. \tag{3.4.2}$$

Therefore, the children belonging to a parent are set to zeros when they are found to be below a certain level of regularity, so that different levels of regularity of an image could be produced by this relationship,

$$R^l_{j,u} f = 2^{\alpha+2}. \tag{3.5}$$

Apart from this, a finer estimation can also be performed with the expense of computations by computing the interscale ratios for each coefficient at the finer scales,

$$R^l_{j,2u_1-u'_1,2u_2-u'} f = \frac{|W^l_j f[u]|}{|W^l_{j-1} f[2u_1 - u'_1, 2u'_2 - u'_2]|}, \quad \text{for } u'_1, u'_2 = 0,1, \tag{3.6}$$

and

$$R^l_{j,u} f = 2^{\alpha}. \tag{3.7}$$

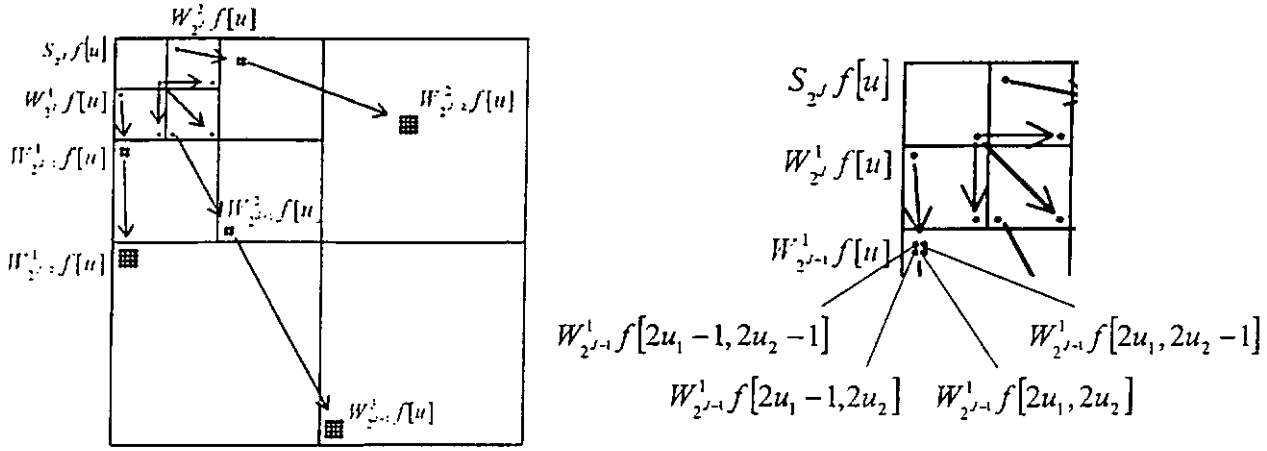


Figure 3.11 The decimated separable wavelet transform and the construction of quad-trees.

It has been shown in [32] that the interscale ratios of the magnitude sums over the COI can give a better estimation of Lipschitz regularity. We further apply this approach to define different levels of regularity for our scalable coding algorithm. Using the interscale ratio of the magnitude sum to estimate the regularity of a signal is especially good for filter bank implementation of the separable wavelet transform with decimation. It is because the maxima points of the wavelet transform may be decimated out and there may have error for the localization of the singular points detected. Therefore, we cannot apply the WTMS approach developed before. To compute the magnitude sum, we need to determine the range, that is the COI, for it. Let us first consider the COI for the magnitude sum (figure 3.12) at v over the undecimated wavelet transform. Suppose ϕ and ψ have compact supports $[-K_1, K_1]$ and $[-K_2, K_2]$ respectively. Denote $u = (u_1, u_2)$ and $v = (v_1, v_2)$. The undecimated COI at v are the following.

$$\text{For } l = 1, \quad COI^1_{j,v} = \left\{(j,u) : |u_1 - v_1| \le K_1 2^j, |u_2 - v_2| \le K_2 2^j\right\}.$$

$$\text{For } l = 2, \quad COI^2_{j,v} = \left\{(j,u) : |u_1 - v_1| \le K_2 2^j, |u_2 - v_2| \le K_1 2^j\right\}.$$

$$\text{For } l = 3, \quad COI^3_{j,v} = \left\{(j,u) : |u_1 - v_1| \le K_2 2^j, |u_2 - v_2| \le K_2 2^j\right\}.$$

When each subband is downsampled by 2^j, its size would be equal to that of the decimated wavelet transform. So we determine the decimated COI for the separable wavelet transform with decimation as the following.

$$\text{For } l=1, \qquad dCOI^1_{j,v}=\left\{(j,u):\left|\frac{u_1}{2^j}-v_1\right|\le K_1,\left|\frac{u_2}{2^j}-v_2\right|\le K_2\right\}.$$

$$\text{For } l=2, \qquad dCOI^2_{j,v}=\left\{(j,u):\left|\frac{u_1}{2^j}-v_1\right|\le K_2,\left|\frac{u_2}{2^j}-v_2\right|\le K_1\right\}.$$

$$\text{For } l=3, \qquad dCOI^3_{j,v}=\left\{(j,u):\left|\frac{u_1}{2^j}-v_1\right|\le K_2,\left|\frac{u_2}{2^j}-v_2\right|\le K_2\right\}.$$

We then compute the magnitude sums,

$$N^l_{j,v}f=\sum_{dCOI^l_{j,v}}\left|W^l_j f[u]\right|, \tag{3.8.1}$$

and

$$N_{J,v}f=\sum_{dCOI^3_{j,v}}\left|S_J f[u]\right|. \tag{3.8.2}$$

Similar to eqn.3.4 and eqn.3.6, we obtain their interscale ratios as,

$$R^l_{j,v}f=\frac{N^l_{j,v}f}{\sum_{v_1'=0}^{1}\sum_{v_2=0}^{1}N^l_{j-1,2v_1-v_2',2v_1-v_2'}f}, \tag{3.9.1}$$

and at the coarsest scale $j=J$,

$$R^l_{J,v}f=\frac{N_{J,v}f}{N^l_{J,v}f}, \tag{3.9.2}$$

or

$$R^l_{j,2v_1-v_1',2v_2-v_2'}f=\frac{N^l_{j+1,v}f}{N^l_{j,2v_1-v_1',2v_2-v_2'}f}, \quad \text{for } v_1', v_2'=0,1. \tag{3.10}$$

By theorem 2.6 and 3.1, we have $N^l_{j,v}f\le A''\left(2^j\right)^{\alpha+1}$, where $A''>0$ is a finite constant. Similar to eqn.3.5 and eqn.3.7, we have the interscale ratios for the magnitude sums over the decimated COI for eqns.3.9 and eqn.3.10 as

$$R^l_{j,v}f=2^{\alpha+3} \tag{3.11}$$

and

$$R^l_{j,v}f=2^{\alpha+1} \tag{3.12}$$

respectively. As a result, we do not need to directly compute the Lipschitz exponents and avoid the errors and ambiguities that occur when tracing the evolution of the wavelet transform coefficient magnitudes at coarser scales.

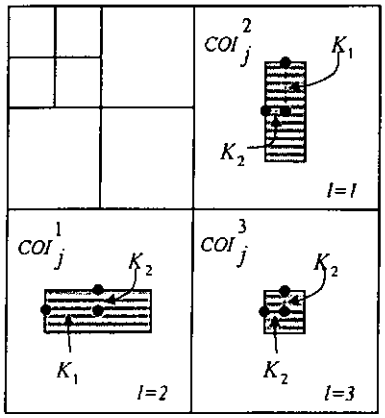


Figure 3.12 The decimated COI of separable discrete wavelet transform.

To develop an efficient image coding algorithm that possesses regularity scalability, the regularity at different parts of an image should first be evaluated. Traditionally, a jointly shiftable wavelet transform [46] is used to represent multiscale signals and perform image analysis, because the transform coefficients are translation invariant in position and orientation. There is also no aliasing effect upon subsamplings and zero interpolation, and the wavelet filters used have compact supports in Fourier domain or finite impulse responses. Since we aim at image compression, it is usually more desirable to obtain a representation of signal information with a minimal degree of correlation and redundancy but a maximal concentration of energy distribution, so that the compression rate is maximized. Therefore, regularity is measured from the wavelet transform coefficients obtained from the transform coding part of the wavelet image coder, even though this is not the

best decomposition for analyzing the features with that set of wavelet filters. In the previous section, we have shown that we can estimate the Lipschitz regularity from the interscale ratio of the transform coefficient magnitude or the magnitude sum over the decimated COI. With the partial ordering of the Lipschitz regularity and magnitude of the wavelet coefficients, the entropy of the ordered coefficients become lower after the first few selections. Then we completely embed the regularity scalable coding into the wavelet image coder so that it becomes simpler to be implemented. In addition, it is less computationally demanding to operate on the decimated wavelet coefficients as fewer coefficients are involved.

Algorithm (selection based on wavelet transform magnitudes)

1. *Obtain the decimated separable wavelet transformed image (eqn.3.1) from the encoder.*
2. $\forall j, l$, *compute and normalize their interscale ratios* $R^l_{j,v}f$ *(eqn.3.4 or eqn.3.6).*
3. *Normalize the magnitudes of the wavelet coefficients in step 1 and the results from step 2 individually. Combine them and compute* $R^l_{j,v}f$ *again after this adjustment.*
4. *Determine* α_i *for* $i = 1,2,...,L$, *where* L *is the regularity levels, so that an even progression of regularity across different regularity levels is obtained.*

Then we set $R^l_{j,v,i}f = 2^{\alpha_i}$ *for each* i.

5. $\forall j, l$, *perform selection on* $W^l_j f[u]$ *by eqn.3.5 or eqn.3.7. That is, for each* i, *we select* $W^l_j f[u]$ *if* $R^l_{j,v}f \le R^l_{j,v,i}f$, *so that those coefficients with larger magnitudes and smaller* $R^l_{j,v}f$ *have the higher priority to be encoded first.*

6. *The selected coefficients are subsequently encoded by the quantizer and the entropy coder to form a regularity and resolution scalable bitstream.*

Algorithm (selection based on the wavelet transform magnitude sum)

1. *Obtain the decimated separable wavelet transformed image (eqn.3.1) from the encoder.*

2. *Compute the magnitude sums using eqn.3.8, where* $1 \le j \le J$ *and* $1 \le l \le 3$.

3. *Compute the interscale ratios* $R^l_{j,v}f$ *of the magnitude sums (eqn.3.10).*

4. *Normalize the magnitudes of the wavelet coefficients (step 1) and the results from step 3 individually. Combine them and compute* $R^l_{j,v}f$ *again after this adjustment.*

5. *Determine* α_i *for* $i = 1,2,...,L$, *where* L *is the total number of regularity levels. Set* $R^l_{j,v,i} = 2^{\alpha_i+1}$ *for each* i *as eqn.3.12.*

6. *Starting at* $j = J$, *for regularity level* i, *select* $W^l_j f[v]$ *if* $R^l_{j,v}f \le R^l_{j,v,i}f$ $\forall j, l$, *so that those coefficients with larger magnitudes and smaller* $R^l_{j,v}f$ *have the higher priority to be encoded first.*

7. *The selected coefficients are subsequently encoded by the quantizer and the entropy coder to form a regularity and resolution scalable bitstream.*

In step 1, we obtain the wavelet coefficients from the image coder. We compute the magnitude sums and also their interscale ratios in step 2 and 3. Wavelet coefficients have large amplitudes where the signal has sharp transitions. The larger the magnitude, the more the energy and the visually important information are contained by the wavelet coefficient. So we combine this measurement with the estimated regularity of the image (step 4). With this combination, we can eventually obtain the optimal arrangement of the encoding order to ensure the lower regularity components of an image are encoded first. In step 2 and 4, we perform normalization (or scaling them to the same range) for an even progression of regularity across the scalability levels. This is because the dynamic ranges of different images are different. We do the combination in the simplest way, that is, a simple matrix multiplication. The regularity levels are defined in step 5. The wavelet coefficients are sorted according to the selection in step 6, so that those coefficients with larger magnitudes and smaller adjusted interscale ratios have a higher priority to be encoded first. This arrangement of bitstream effectively increases the coding efficiency of the regularity scalable coded images at different resolutions because the entropy of the first few regularity levels of wavelet coefficients are lowered after this utilization.

entropy of the first few regularity levels of wavelet coefficients are lowered after this utilization.

We also propose a possible implementation of the regularity scalable wavelet image coder as shown in figure 3.13. It shows how the scalable coding algorithm can be embedded in a zerotree wavelet image coder. In summary, the regularity of a recovered image depends on the bit rate budget allowed, which is also decided by criteria such as the resolution required, the fixed or varying quantization steps used, the precision of the wavelet coefficients required or the number of significant bits used.

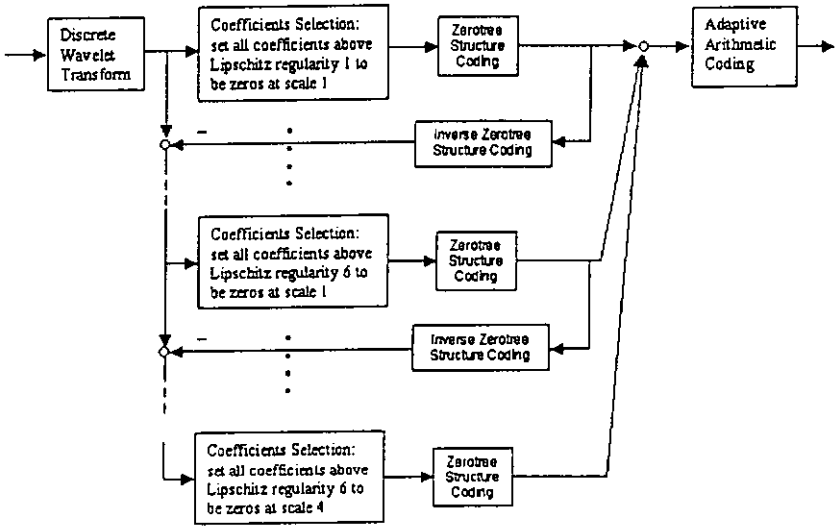


Figure 3.13 The combination of the proposed algorithm with a wavelet zerotree image encoder.

3.4 Results

We combine the proposed algorithm to MPEG-4 still image coder with the popular Daubechies 9/3 biorthogonal filters ($K_1 = 4, K_2 = 1$) (appendix IV). The proposed algorithm introduces regularity scalability in addition to the resolution scalability with no change in bitstream format. This is an important advantage of the proposed approach. Basically it can work with any kind of scalable coding techniques without changing the bitstream format. Well-known USC test images "Lena" and "Peppers" are used as test images. Masks shown in figure 3.14 indicate the positions of the selected wavelet coefficients according to different regularity levels determined by the interscale ratios of magnitudes and magnitude sums. The selected coefficients are encoded and transmitted first. Note that the lower regularity wavelet coefficients associated with the edges and textures are selected first, followed by the higher regularity wavelet coefficients which correspond to the smoother areas or surfaces of the objects.

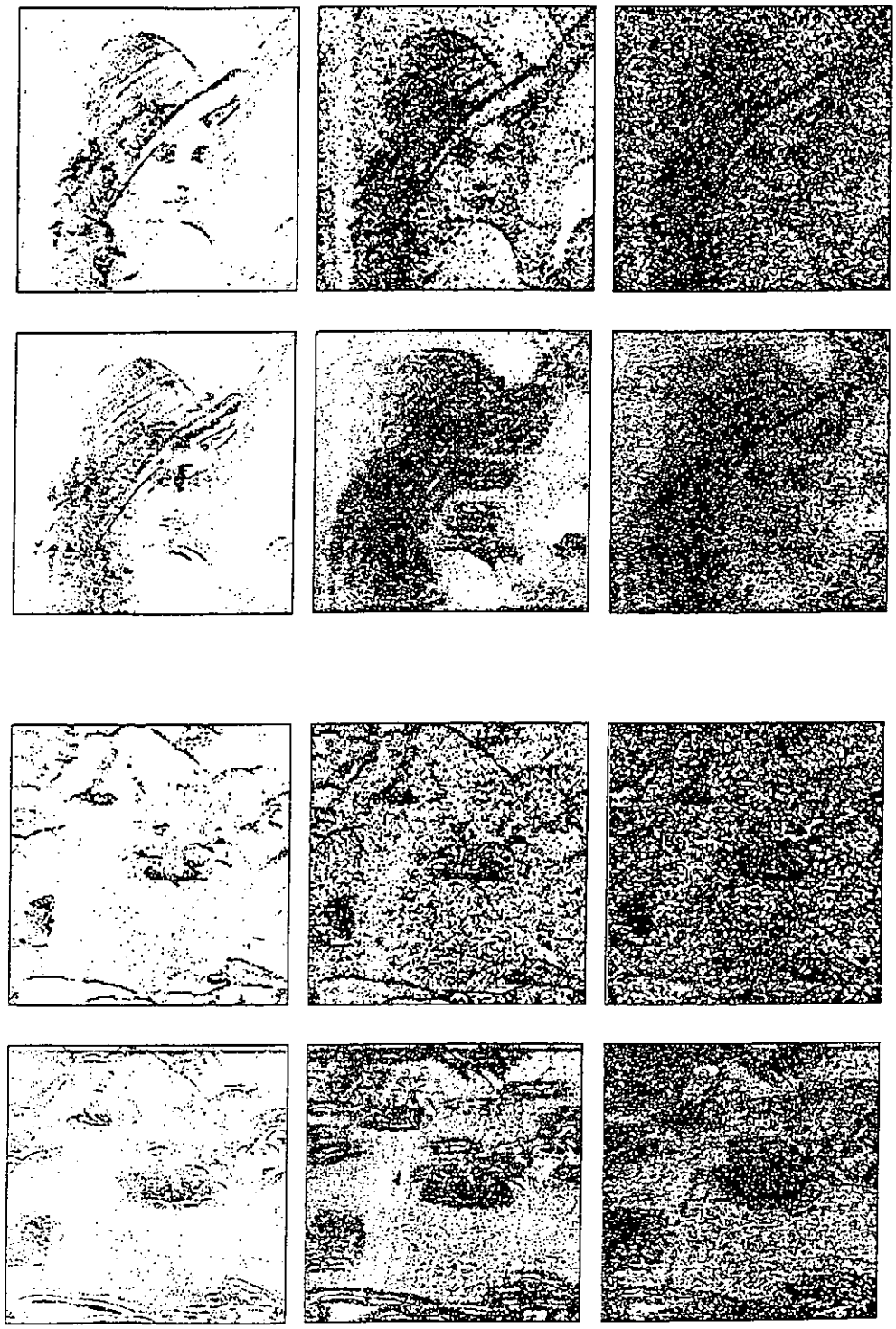


Figure 3.14(a) Masks (l=1 subbands) of the selected wavelet coefficients by different regularity levels from the selection based on the interscale ratios of wavelet transform magnitudes (top), and wavelet transform magnitude sums (bottom).

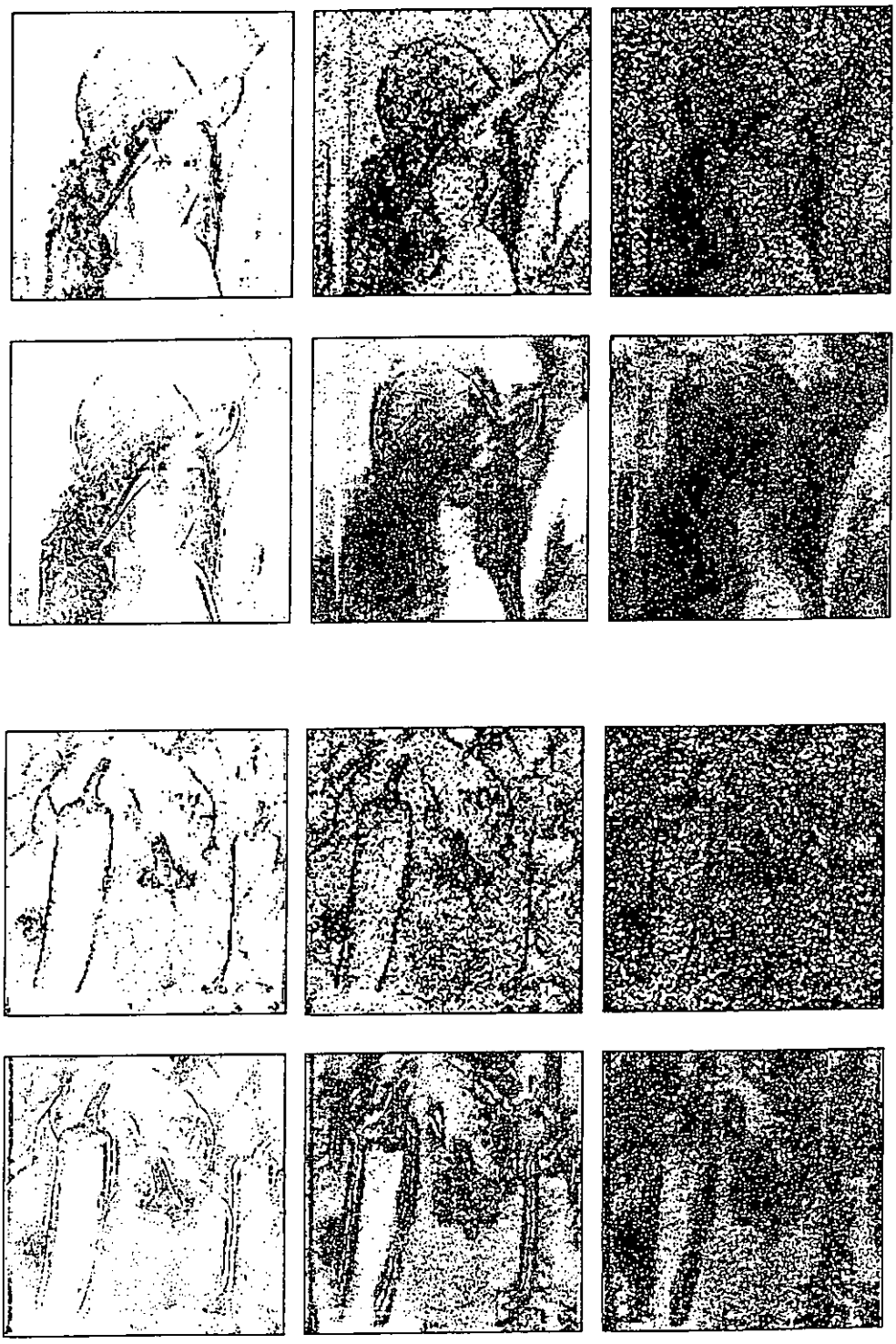


Figure 3.14(b) Masks (l=2 subbands) of the selected wavelet coefficients by different regularity levels from the selection based on the interscale ratios of wavelet transform magnitudes (top), and wavelet transform magnitude sums (bottom).

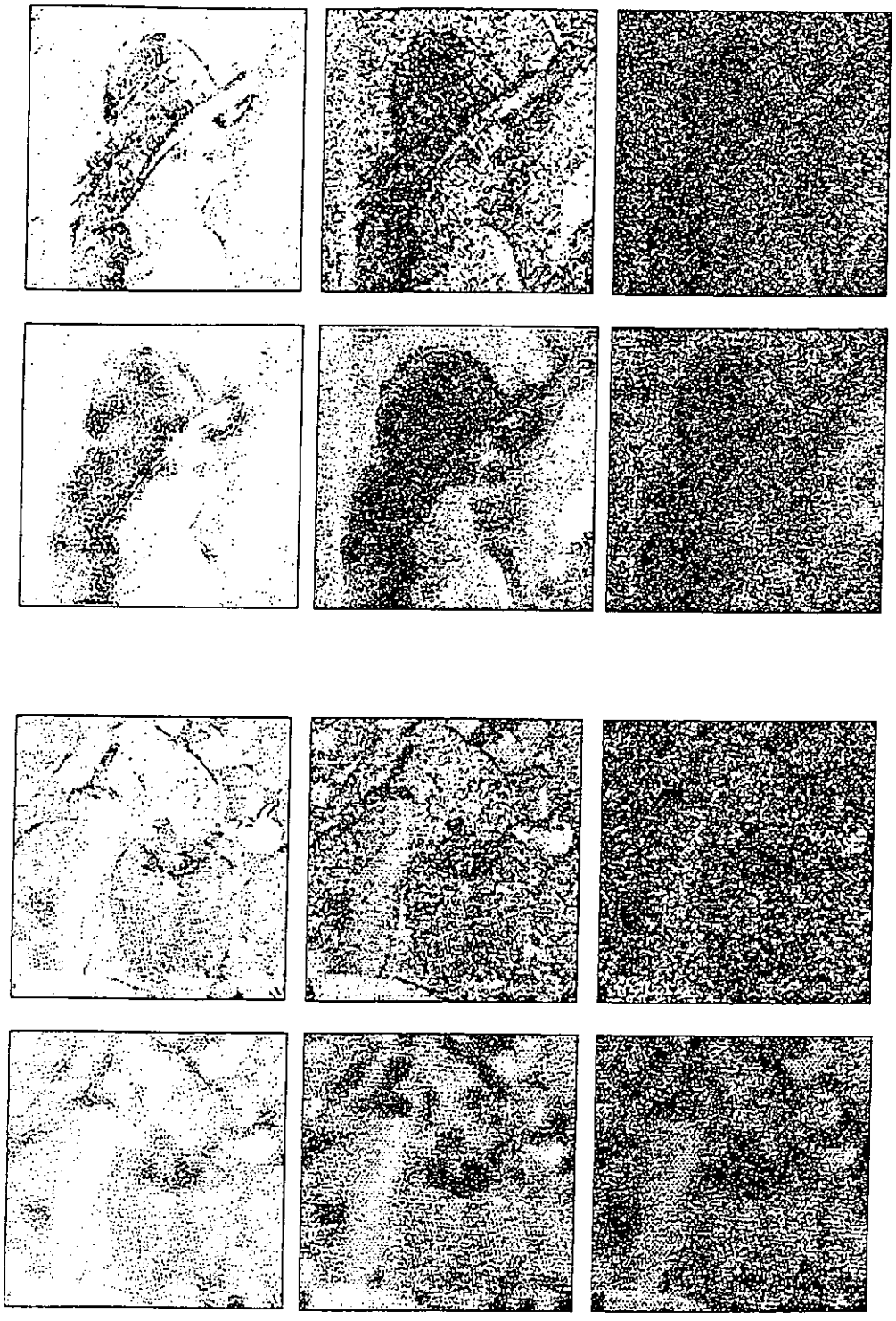
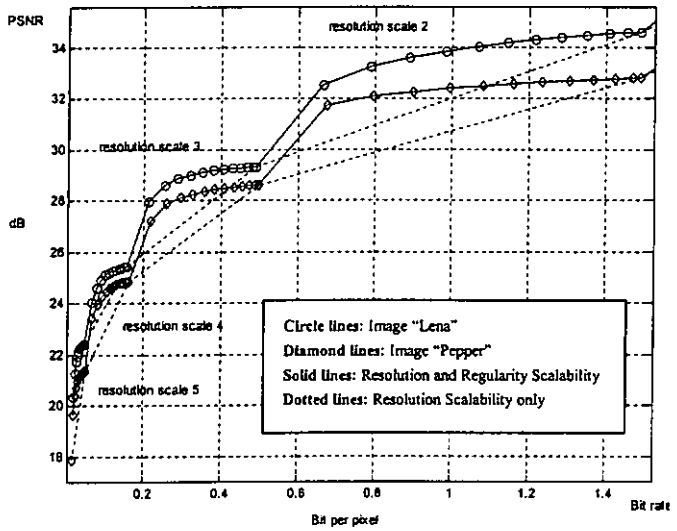
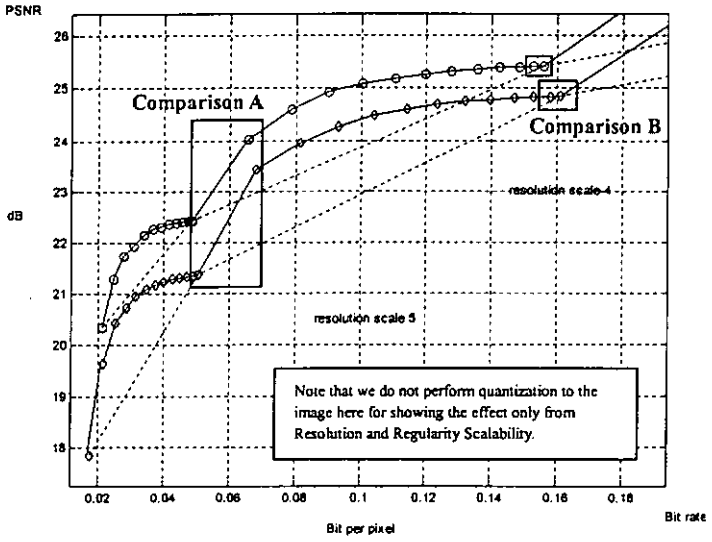


Figure 3.14(c) Masks (l=3 subbands) of the selected wavelet coefficients by different regularity levels from the selection based on the interscale ratios of wavelet transform magnitudes (top), and wavelet transform magnitude sums (bottom).

In figure 3.15, 3.16 and 3.17, the recovered images from our proposed algorithm are compared with the recovered images from the original coding algorithm with resolution scalability only. From the curves (figure 3.15), we can see the rate of change of PSNR against the bit rate at each resolution scale. The overall PSNR for all resolution scales of our scalable coding algorithm outperforms significantly the original coding algorithm with resolution scalability only.



(a)



(b)

Figure 3.15 Performances of Resolution Scalability VS. Resolution and Regularity Scalability. (a) at resolution scales 2-5, (b) at resolution scales 4-5.

Table 3.1 shows the quantitative results (comparison A). For example, when the quantization step size equals to 6, we can see that up to 34.98% reduction in bit rate can be obtained with almost the same quality (less than 2.53% reduction in PSNR) at the resolution scale 2 for encoding the image "Lena". Figure 3.16 and 3.17 shows the qualitative comparison results (comparison B). All the recovered images are scaled to the same size in the illustration. Since the selected coefficients are encoded and transmitted first, as shown in the figures, lower regularity components such as the edges and textures are enhanced significantly with a small addition of bit rate.

	Resolution Scalable		Regularity and resolution Scalable		Percentage Change	
Qstep size	PSNR (dB)	Bit rates (bpp)	PSNR (dB)	Bit rates (bpp)	PSNR (%)	Bit rates (%)
2	34.5795	1.49017	33.5911	0.83072	-2.86	-44.25
4	34.5229	1.14035	33.5762	0.67340	-2.74	-40.95
6	34.4211	0.89929	33.5495	0.58475	-2.53	-34.98
8	34.3012	0.73837	33.5113	0.52167	-2.30	-29.35
32	32.5595	0.25671	32.4216	0.24667	-0.42	-3.91
64	.30.4063	0.13934	30.4040	0.13931	-0.01	-0.02

Table 3.1 Effect of quantization to the regularity scalable coding at resolution scale 2.

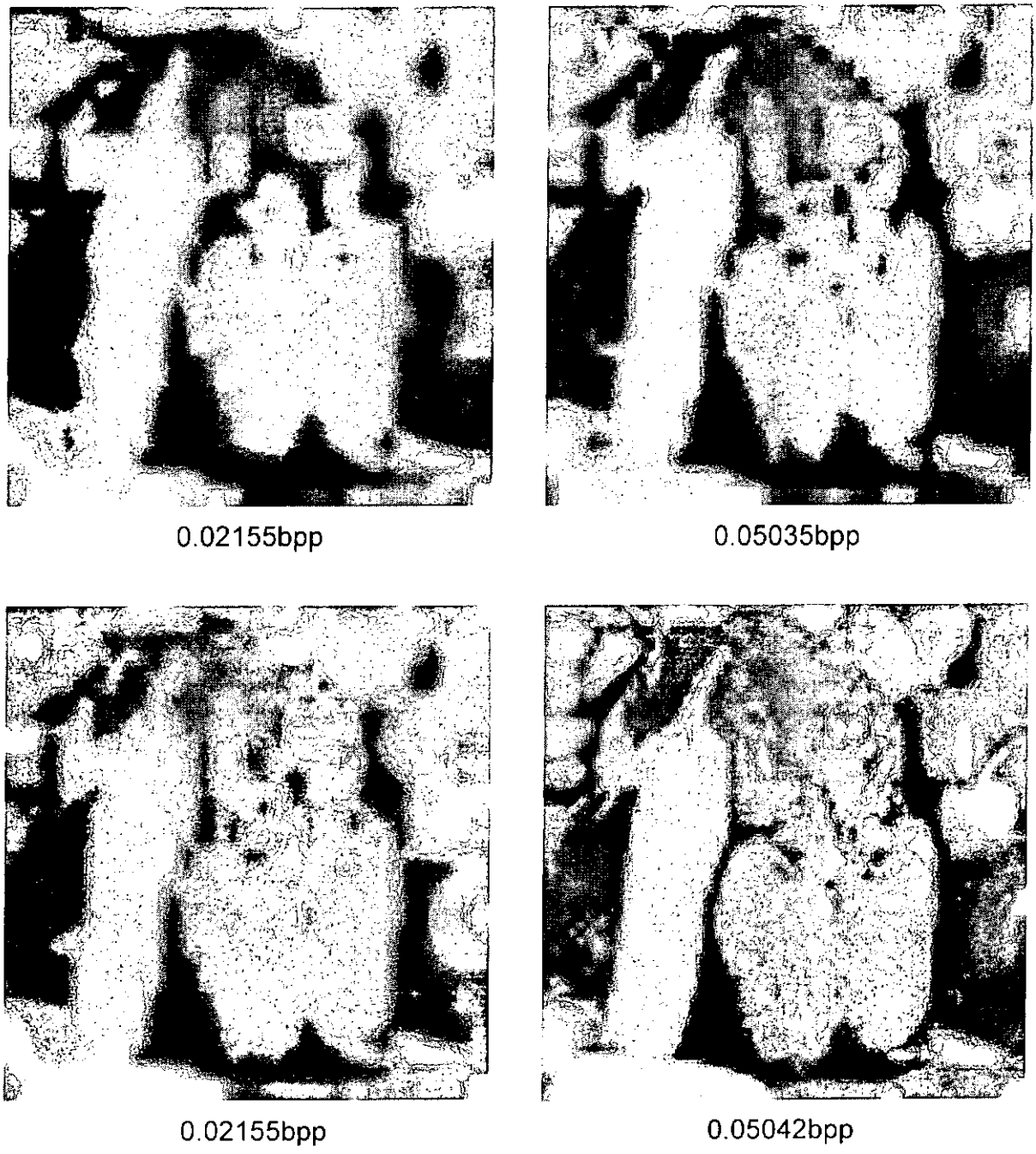


Figure 3.16(a) Reconstructed images "Pepper". 1st row: Resolution Scalability only, 2nd row: Resolution and Regularity Scalability; 1st column: at resolution scale 5, 2nd column: at resolution scale 4.

Figure 3.16(b) Reconstructed images "Pepper". 1st row: Resolution Scalability only, 2nd row: Resolution and Regularity Scalability; 1st column: at resolution scale 3, 2nd column: at resolution scale 2.

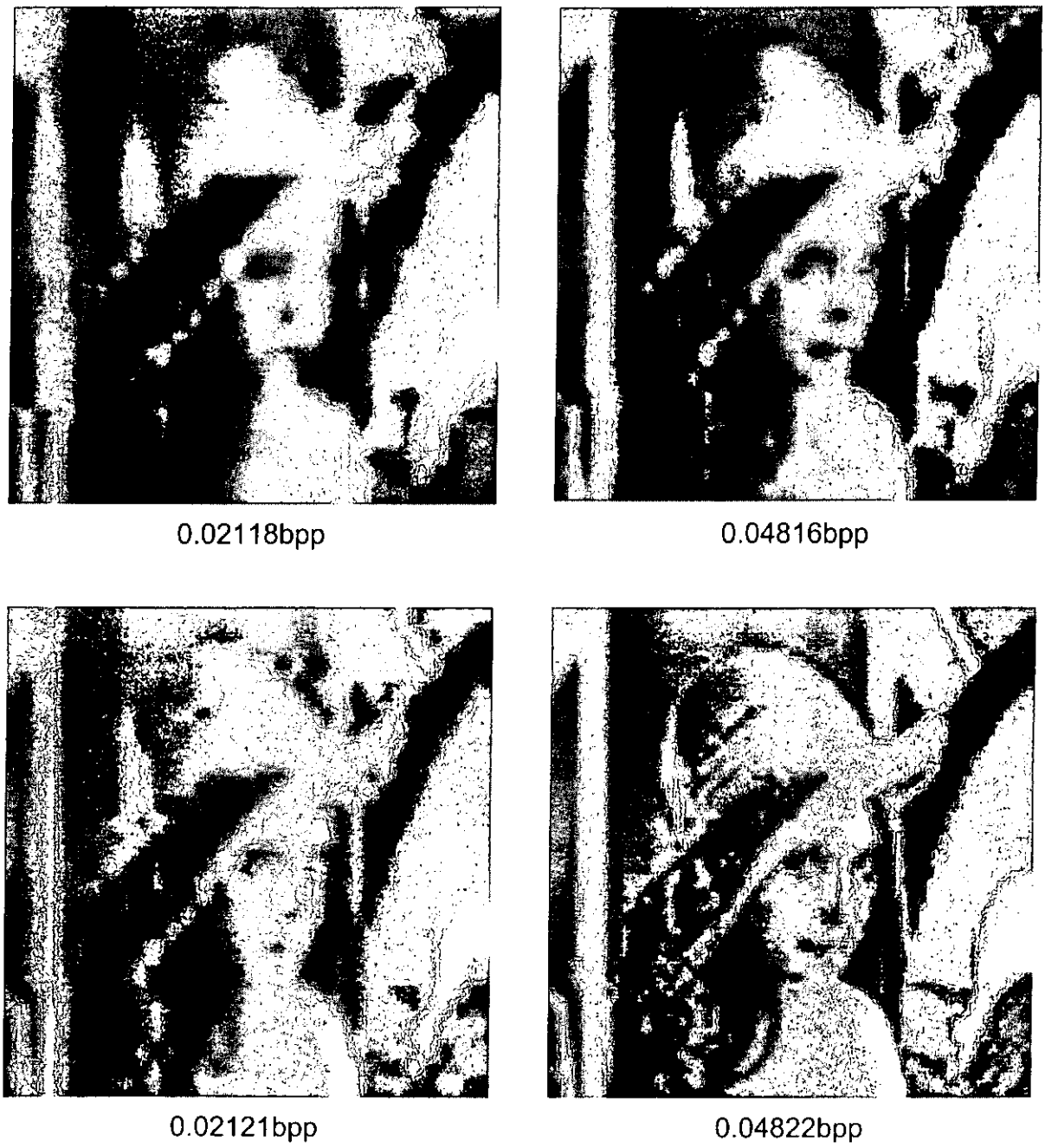


Figure 3.17(a) Reconstructed images "Lena". 1st row: Resolution Scalability only, 2nd row: Resolution and Regularity Scalability; 1st column: at resolution scale 5, 2nd column: at resolution scale 4.

Figure 3.17(b) Reconstructed images "Lena". 1st row: Resolution Scalability only, 2nd row: Resolution and Regularity Scalability; 1st column: at resolution scale 3, 2nd column: at resolution scale 2.

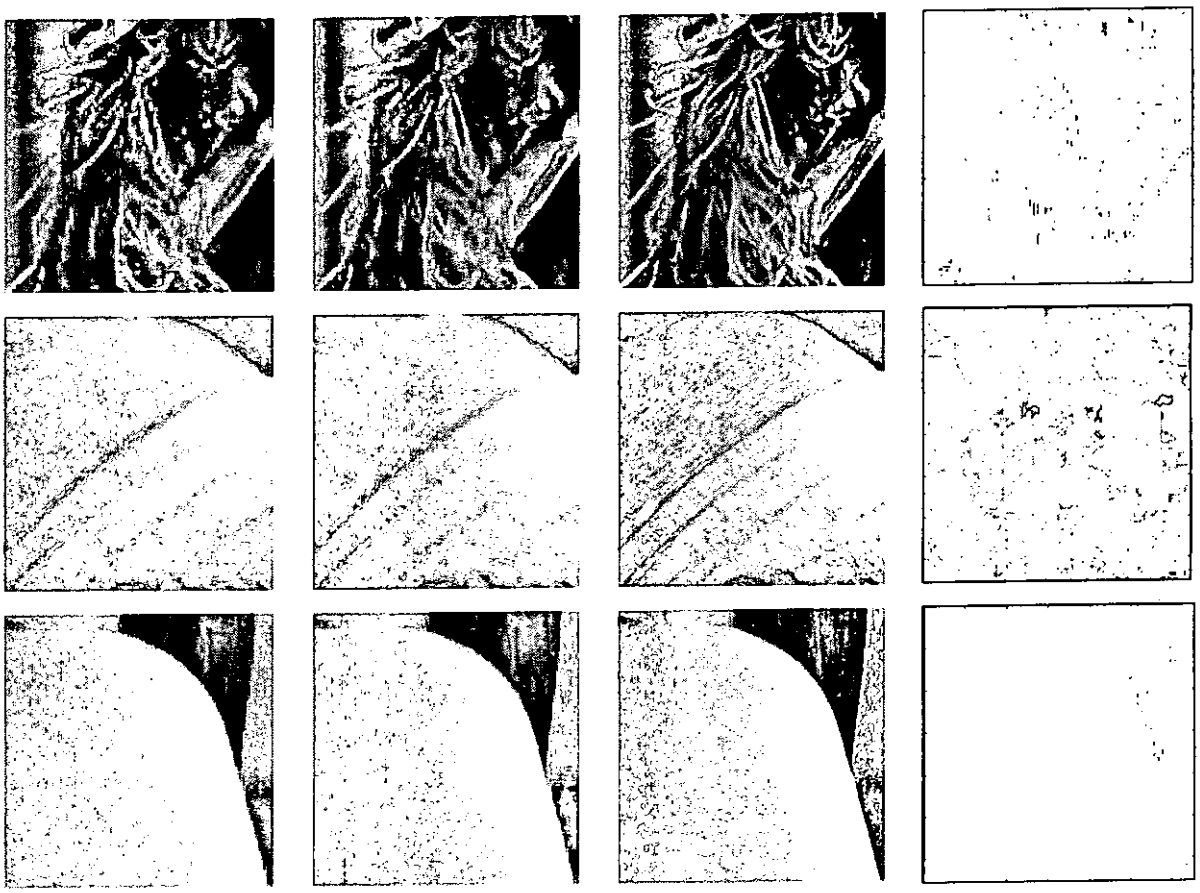
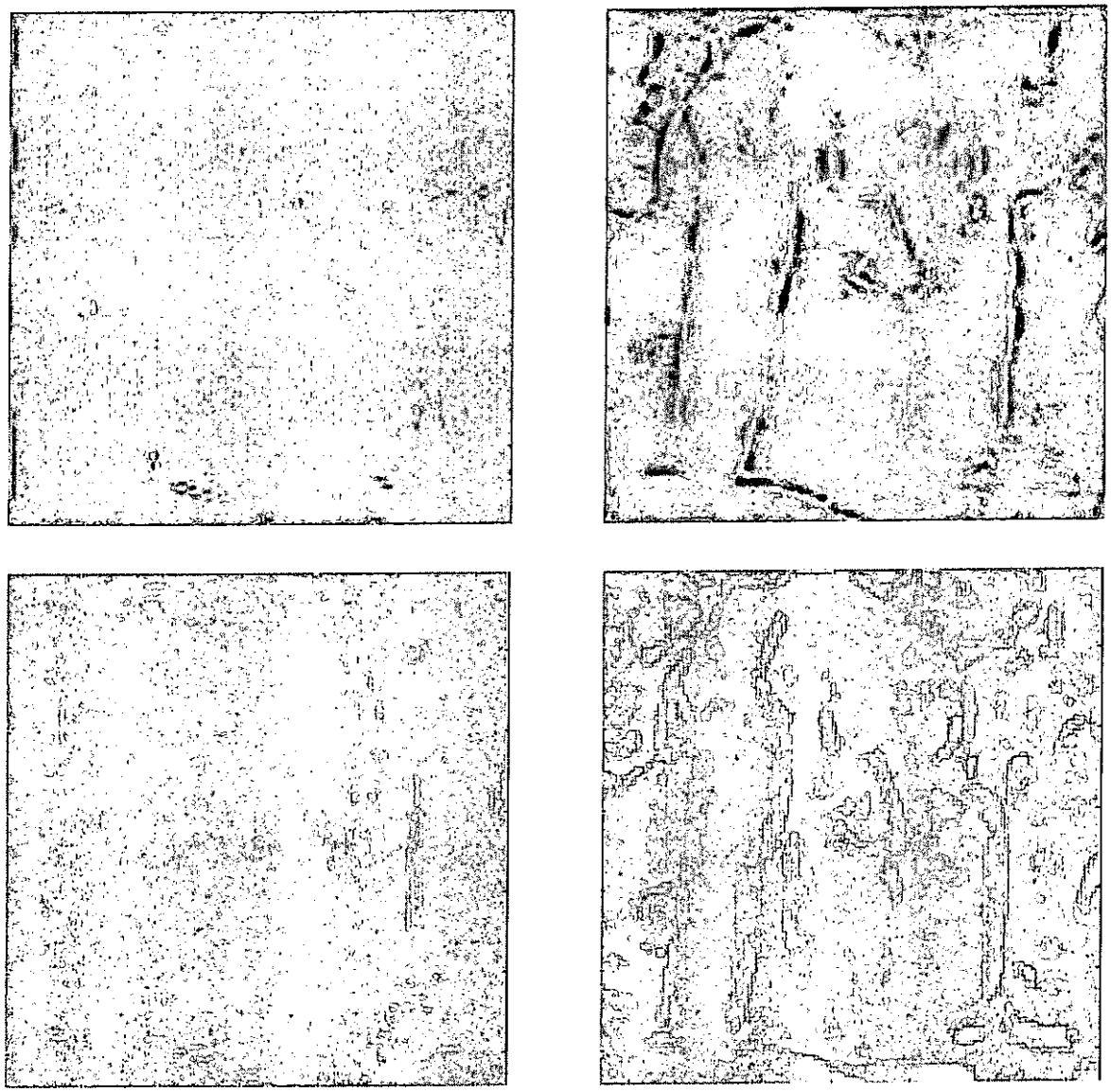


Figure 3.18 Different parts of the reconstructed image "Lena" at resolution scale 2 with the highest precision. (a) 1st column: resolution scalable only (34.5795dB, 1.49017bpp); (b) 2nd column: regularity and resolution scalable (33.5911dB, 0.83072bpp); (c) 3rd column: original image at 3.25668bpp; 4th column: differences of 1st and 2nd column images.

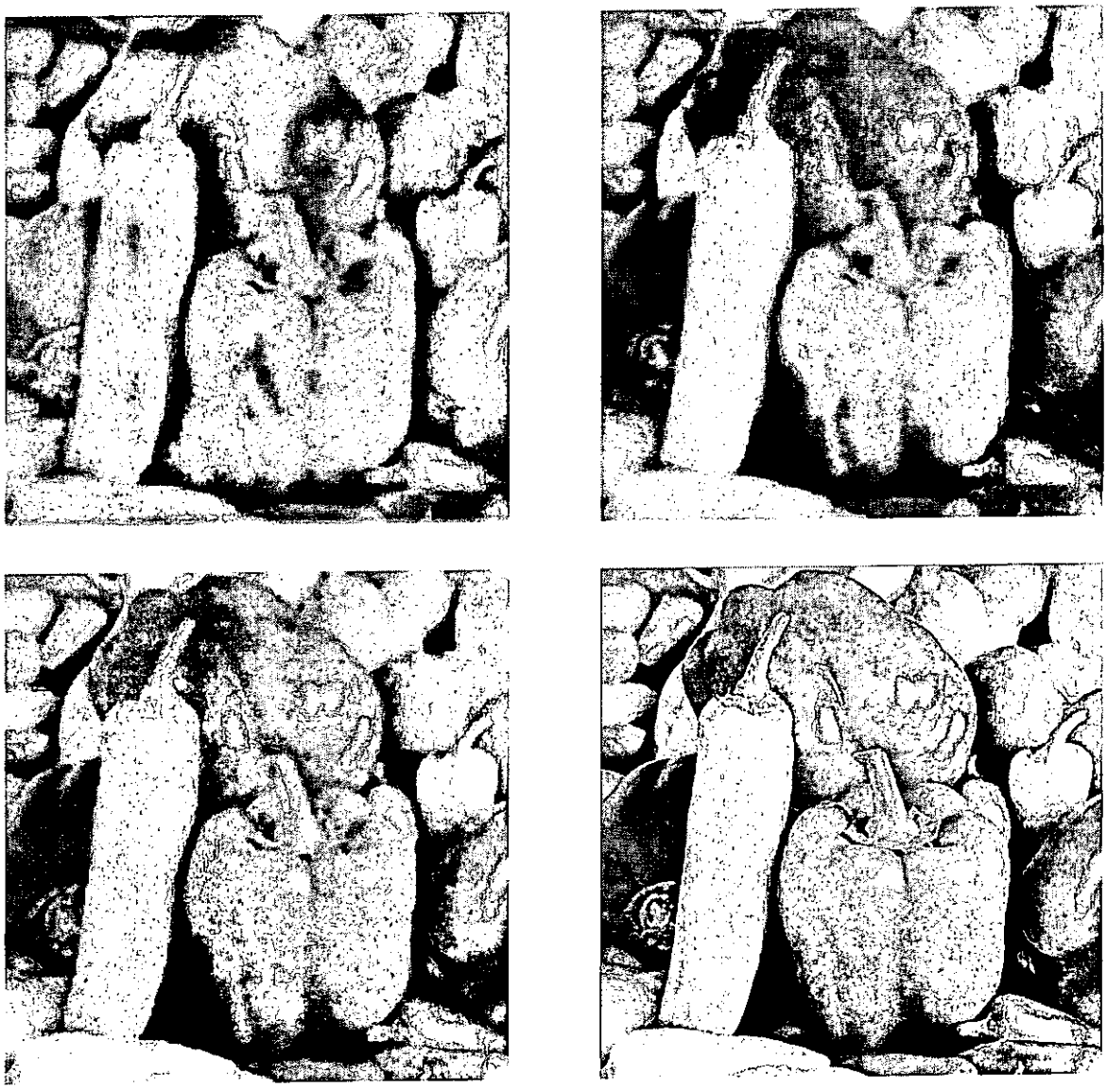
Recall that the embedded coding is initiated with the largest quantization step which is progressively decreased. The coding precision increases as the number of quantization refinement increases. Therefore, to visualize the best qualitative and quantitative improvement from the regularity scalability, we preserve the highest precision by not performing quantization for the results in figure 3.15 and 3.18.

Recall that [91] suggested encoding the higher resolution subbands first, so that the visually important edges are decoded and displayed first, while it was traditionally suggested that the lowest resolution subband should be encoded first so that the gray levels can be displayed at all time for a visually better decoded image. Here we adopt our proposed algorithm into these two transmission schemes. We use the popular 9/3 biorthogonal wavelet filters with a five-level transform. Results of the well-known USC images "Pepper" are presented. For the transmission scheme based on [91], which is shown in figure 3.19, it is worth to note that the skeleton shapes of the images are significantly enhanced at very low bit rate. We can see that the overall quality of the decoded images starting at the same bit rate is also improved. For the traditional transmission scheme with the surface-oriented scalability [94], we found that the overall quality, especially the boundary shapes and edges of the decoded images, are improved with our proposed algorithm.



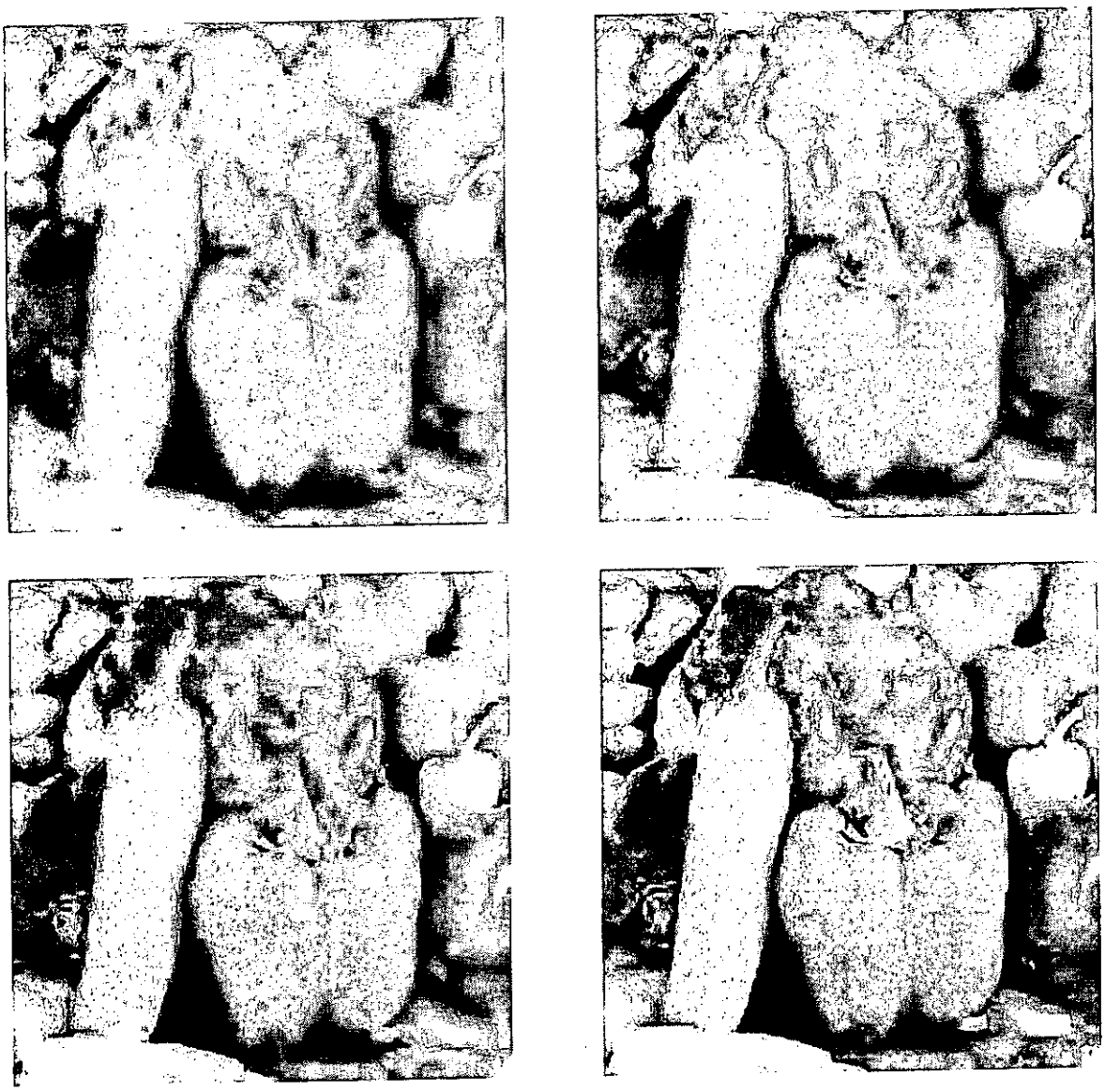
(i) (ii)

Figure 3.19(a) Comparison of the recovered images "Pepper" by our proposed algorithm [86] and the edge-oriented progressive coding algorithm [91]. (i) top: 0.0120bpp from [91]; bottom: 0.0120bpp from our proposed algorithm [86], (ii) top: 0.2227bpp from [91]; bottom: 0.2203bpp from our proposed algorithm [86].



(iii) (iv)

Figure 3.19(b) Comparison of the recovered images "Pepper" by our proposed algorithm [86] and the edge-oriented progressive coding algorithm [91]. (iii) top: 0.3004bpp from [91]; bottom: 0.2990bpp from our proposed algorithm [86], (iv) top: 0.3274bpp from [91]; bottom: 0.3235bpp from our proposed algorithm [86].



(i) (ii)

Figure 3.20(a) Comparison of the recovered images "Pepper" by our proposed algorithm [86] and the surface-oriented scalable coding algorithm [94]. (i) top: 0.0367bpp from [94]; bottom: 0.03738bpp from our proposed algorithm [86]. (ii) top: 0.0524bpp from [94]; bottom: 0.05368bpp from our proposed algorithm [86].

(iii) (iv)

Figure 3.20(b) Comparison of the recovered images "Pepper" by our proposed algorithm [86] and the surface-oriented scalable coding algorithm [94]. (iii) top: 0.1442bpp from [94]; bottom: 0.1443bpp from our proposed algorithm [86], (iv) top: 0.2808bpp from [94]; bottom: 0.2822bpp from our proposed algorithm [86].

3.5 Summary

In this chapter, an adaptive regularity scalable wavelet image coding algorithm was presented. Based on the Lipschitz regularity condition derived for the separable wavelet transform, regularity measurement and singularity detection can be performed in image coding algorithm. We first give a regularity scalable representation of an image by selecting the wavelet coefficients according to their interscale ratios. WTMS over the COI can provide a better regularity estimation, but we cannot directly apply it to determine the scalable representation. So we select the wavelet coefficients in various extents according to the interscale ratios of their magnitude sums over the decimated COI. A wavelet image decomposition provides a hierarchical data structure, and this selection utilizes the tree structured data organization. With partial ordering of the wavelet coefficient magnitudes, the entropy of the selected coefficients at the first few regularity levels are lowered. This accounts for why the regularity scalable coding algorithm can provide us a higher image compression capability.

From the simulation results and the comparison with several scalable coding techniques, we conclude that our regularity scalability can provide a better visual perception of an image at very low bit rate. Moreover, the algorithm operates on the decimated wavelet coefficients and hence much less computation is needed than operating on the undecimated coefficients. The proposed algorithm is adopted in

MPEG-4 still image codec and benefited by its high coding efficiency. It can also work with other kinds of scalabilities such as the resolution scalability, without changing the bitstream format. We found that the proposed algorithm outperforms various existing scalable coding algorithms in terms of visual perception, implementation complexity and coding efficiency. In addition, it can possess good noise robustness by appropriately adjusting the threshold of the first regularity level, which behave as the adaptive denoising algorithms in [32,89,6], to reject wavelet coefficients with negative Lipschitz exponents which correspond to noise.

This scalable coding algorithm would be very useful in image browsing and retrieval applications. For image database browsing, a user may want to search through the database to find an image with particular features but minimize the number of bits transmitted for faster response. With the proposed scalable coding algorithm, which is based on Lipschitz regularity, an image becomes easier to be recognized at low bit rates and any of the important features of an image can be visualized first at a certain bit rate.

Chapter 4

Multiwavelet Denoising using Singularity Detection

4.1 Introduction

Multiwavelets have been proposed as a generalization of the traditional scalar wavelet, with multiresolution analysis of multiplicity r. It has been proposed that multiwavelet bases should be better in wavelet applications than other wavelet bases [48]. It is because there are limitations for the single wavelet decomposition. First of all, it cannot offer simultaneous orthogonality, symmetry and compact support together with high regularity (or approximation order greater than one) [66,67]. Among the single wavelets only the Haar wavelet fulfils the above properties, but it is discontinuous in the spatial domain. Multiwavelet systems based on two scaling functions and two wavelets allow these properties simultaneously. Therefore, the design of multiwavelet filters is more flexible.

Multiwavelet transform has been applied to image compression and denoising in the past few years. Experimental results [18,49,92,99] showed that the multiwavelets generally outperform the single wavelet in signal and image denoising using the

thresholding approach. It is because the multiwavelets can offer higher vanishing moments with shorter filter supports. The number of high amplitude wavelet coefficients created by a brutal transition like an edge is proportional to the size of the supports of the filters. For a more accurate localization of singularities, the number of high amplitude wavelet coefficients produced should be as small as possible, and so the supports of the filters should be as short as possible. Moreover, the more the vanishing moments, the smaller the coefficients can be produced at fine scales. Therefore, the multiwavelet coefficients that belong to the noise component can be more easily distinguished at fine scales.

In this chapter, we study the approaches for signal and image denoising using singularity detection with the multiwavelets, as a work further extended from the regularity scalable coding. We investigate the evolution of the multiwavelet coefficients across scales from their interscale ratios. Then we perform the magnitude sums over the COI and estimate the regularity from their interscale ratios, as it can be verified that the COI can be more clearly determined than in the single wavelet transform, especially at lower scales. To illustrate the advance of the multiwavelet regularity estimation, we propose a signal denoising based on this. The denoising algorithm treats the individual coefficients or the coefficients vectors as a whole entity. Compared to the single wavelet case [32,88], improved results for singularity

detection and signal denoising can be achieved for white Gaussian noise and impulsive noise.

The organization of this chapter is the following. We first review the current signal and image denoising techniques using the single wavelet and multiwavelet transform. We then develop the Lipschitz condition for the multiwavelet transform. The denoising algorithm is presented next, followed by the results and the summary.

4.2 The present works

Signal and image denoising using the wavelet transform began to receive much attention since 1992, when Mallat proposed to use the WTMM to estimate the local Lipschitz exponent of a signal by tracing the evolution of its WTMM across scales [39] (Please refer to section 2.2.4 for the detail.). However, irregular sampling nature complicates the reconstruction process. On the other hand, signal and image denoising using thresholding techniques was pioneered by Donoho and Johnstone in 1994 and 1995 [13,14,15,16]. Soft or hard thresholding is performed to the detail discrete wavelet coefficients of a noisy signal. Although it offers the advantages of smoothness and adaptation, it produces visual artifacts. Translation-invariant denoising scheme [65] was proposed to suppress such artifacts by averaging the denoised signals. Later in 1997, a signal and image denoising algorithm which combines the approaches of Mallat and Donoho without their disadvantages, was proposed by Hsung *et. al* [32,88]. Thresholding is performed to the detail wavelet coefficients according to the interscale ratio and difference of the magnitude sum of the wavelet coefficients over the COI and the denoised signal can be obtained by simple inverse wavelet transform.

Since the multiwavelet transform has some nice properties that cannot be obtained from the single wavelet transform, and it offers higher flexibility in designing the filters. Strela *et. al* suggested a way to construct biorthogonal multiscaling functions

and applied it to signal and image denoising [49,99]. Downie and Silverman fully utilized the multiple wavelet transform for signal and image denoising by bivariate thresholding [18]. Several threshold estimators for the single wavelet transform are experimented by Ling for the multiwavelet transform [93]. We can see that denoising algorithms and threshold estimators using the single wavelet transform can be extended to multiwavelet systems and improved results can be obtained.

4.2.1 Thresholding estimation

Denoising problem is actually the estimation of signals in additive noise, which is optimized by finding a representation that discriminates the signal from the noise. Suppose that a signal of interest f has been corrupted by noise, so that we observe a signal g: $g[n] = f[n] + \sigma z[n]$, $n = 0,1,\cdots,N-1$, where $z[n]$ is a unit-variance, zero-mean white Gaussian noise. We consider estimators computed with an orthonormal basis $\mathrm{B} = \{h_m\}_{0 \le m < N}$, and the noisy data is decomposed in B. The inner product with h_m gives $g_{\mathrm{B}}[m] = \langle g, h_m \rangle$, $f_{\mathrm{B}}[m] = \langle f, h_m \rangle$ and $z_{\mathrm{B}}[m] = \langle z, h_m \rangle$ and so $g_{\mathrm{B}}[n] = f_{\mathrm{B}}[n] + \sigma z_{\mathrm{B}}[n]$.

An estimation is calculated by an operator that attenuates the noise while preserving the signal. There are two ways to perform thresholding estimation, the *hard* and *soft* thresholding. A hard thresholding is implemented with

$$\rho_\lambda(x) = y = \begin{cases} x & \text{if } |x| > \lambda \\ 0 & \text{if } |x| \le \lambda \end{cases}. \tag{4.1}$$

A soft thresholding is implemented with

$$\rho_\lambda(x) = y = \begin{cases} x-\lambda & \text{if } x \ge \lambda \\ x+\lambda & \text{if } x \le -\lambda \\ 0 & \text{if } |x| \le \lambda \end{cases}. \tag{4.2}$$

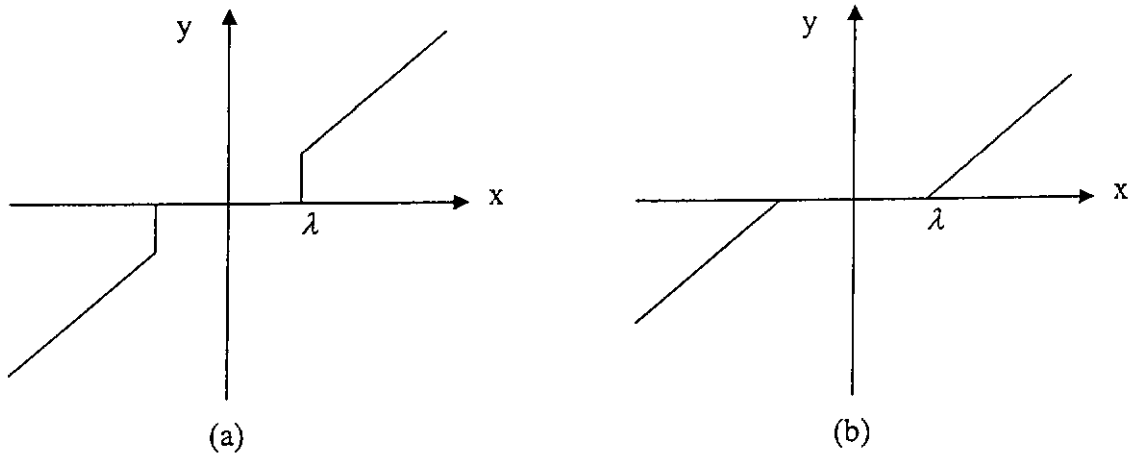


Figure 4.1 (a) Hard thresholding (b) Soft thresholding.

The hard thresholding will kill all the transformed coefficients whose magnitudes are less than the threshold to zero while keeping the remaining coefficients unchanged. The soft thresholding kills the smaller coefficients as well. However, all the coefficients whose magnitudes are greater than the threshold will be reduced by the amount of the threshold. The risk (or the mean square error) of hard thresholding is

$$r(f,\lambda) = \sum_{m=0}^{N-1} E\left\{ \left| f_B[m] - \rho_\lambda(g_B[m]) \right|^2 \right\}, \tag{4.3}$$

which depends on the threshold λ used.

In addition, the *semi-soft* thresholding, which was proposed in [103], generalizes the hard and soft functions by using two thresholds, including both the hard and soft as

the special case.

4.2.2 Threshold refinement

Suppose $\lambda = \sigma\sqrt{2\log_e N}$, to reduce the thresholding risk, we can actually choose a threshold which is smaller than λ. Then the threshold is adapted to the data and it can be calculated by minimizing an estimation of the risk. Improvement of thresholding estimators can be obtained with a translation invariant algorithm.

SURE Threshold

To study the impact of the threshold on the risk, we denote by $r(f,\lambda)$ the risk of a soft thresholding estimator calculated with a threshold λ. An estimate $\tilde{r}(f,\lambda)$ of $r(f,\lambda)$ is calculated from the noisy data g, and λ is optimized by minimizing $\tilde{r}(f,\lambda)$.

To estimate the risk $r(f,\lambda)$, observe that if $|g_B[m]| < \lambda$ then the soft thresholding sets this coefficient to zero, which produces a risk equal to $|f_B[m]|^2$. Since $E\{|g_B[m]|^2\} = |f_B[m]|^2 + \sigma^2$, one can estimate $|f_B[m]|^2$ with $|g_B[m]|^2 - \sigma^2$. If $|g_B[m]| \geq \lambda$, the soft thresholding substracts λ from the amplitude of $g_B[m]$. The expected risk is the sum of the noise energy plus the bias introduced by the reduction of the amplitude of $g_B[m]$ by λ. It is estimated by $\sigma^2 + \lambda^2$. The resulting estimator of $r(f,\lambda)$ is

$$\tilde{r}(f,\lambda)=\sum_{m=0}^{N-1}\left(\left|g_B[m]\right|^2\right) \tag{4.4}$$

with $$\Phi(u)=\begin{cases}u-\sigma^2 & if\ u\le\lambda^2\\ \sigma^2+\lambda^2 & if\ u>\lambda\end{cases}.$$

To find the $\tilde{\lambda}$ that minimizes the Stein's unbiased risk estimation (SURE) $\tilde{r}(f,\lambda)$, the N data coefficients $g_B[m]$ are sorted in decreasing amplitude order with $O(N\log_2 N)$ operations. Let $g_B^r[k]=g_B[m_k]$ be the coefficient of rank k : $g_B^r[k]\ge g_B^r[k+1]$ for $1\le k<N$. Let l be the index such that $\left|g_B^r[l]\right|\le\lambda<\left|g_B^r[l+1]\right|$. We can rewrite eqn.4.4 as

$$\tilde{r}(f,\lambda)=\sum_{k=l}^{N}\left|g_B^r[k]\right|^2-(N-1)\sigma^2+l\left(\sigma^2+\lambda^2\right) \tag{4.5}$$

To minimize $\tilde{r}(f,\lambda)$ we must choose $\lambda=\left|g_B^r[l]\right|$ because $r(f,\lambda)$ is increasing in λ. To find the $\tilde{\lambda}$ that minimizes $\tilde{r}(f,\lambda)$, it is therefore sufficient to compare the N possible values $\left\{\left|g_B^r[l]\right|\right\}_{1\le k\le N}$, that requires $O(N)$ operations if we progressively recomputed the eqn.4.5. The calculation of $\tilde{\lambda}$ is thus performed with $O(N\log_2 N)$ operations.

When signal energy is small relative to the noise energy, that is $\|f\|^2<<E\left\{\|\sigma Z\|^2\right\}=N\sigma^2$. In this case, one must impose $\lambda=\sigma\sqrt{2\log_e N}$ in order to remove all the noise. Since $E\left\{\|g\|^2\right\}=\|f\|^2+N\sigma^2$, we estimate $\|f\|^2$ with $\|g\|^2-N\sigma^2$ and compare this value with a minimum energy level $\varepsilon_N=\sigma^2N^{1/2}(\log_e N)^{3/2}$. The resulting SURE threshold is

$$T = \begin{cases} \sigma\sqrt{2\log_e N} & \text{if } \|g\|^2 - N\sigma^2 \le \varepsilon_N \\ \tilde{T} & \text{if } \|g\|^2 - N\sigma^2 > \varepsilon_N \end{cases} \quad (4.6)$$

Translation Invariant Thresholding

An improved thresholding estimator is calculated by averaging estimators for translated versions of the signal. The translation invariant algorithm of Coifman and Donoho [65,102] estimates all translations of f and averages them after a reverse translation. For all $0 \le p \le N$, the estimator $\tilde{F}^p$ of f^p is computed by thresholding the translated data $g^p[n] = f[n-p]$ as $\tilde{F}^p = \sum_{m=0}^{N-1} \rho_\lambda\left(g_B^p[m]\right) g_m$. The translation invariant estimator is obtained by shifting back and averaging these estimates as

$$\tilde{F}[n] = \frac{1}{N}\sum_{p=0}^{N-1} \tilde{F}^p[n+p]. \quad (4.7)$$

In general, this requires N times more calculations than for a standard thresholding estimator.

Other variation of the threshold refinement such as *BayesShrink*, which advocates soft shrinkage in a certain Bayesian framework and outperforms *SureShrink* estimation [16] in the context of denoising image, can be found in [7]. Cross-validation is also applied to determine the threshold [33,42].

4.2.3 Wavelet thresholding

Thresholding wavelet coefficients is equivalent to estimating the signal by averaging it with a kernel that is locally adapted to the signal regularity [61]. A filter bank of conjugate mirror filters decomposes a discrete signal in a discrete orthogonal wavelet basis. The discrete wavelets $\psi_{j,m}[n]=\psi_j[n-N2^j m]$ are translated modulo modifications near the boundaries. The support of the signal is normalized to $[0,1]$ and has N samples spaced by N^{-1}. The scale parameter 2^j thus varies from $2^L = N^{-1}$ up to $2^J < 1$, $B=\left[\{\psi_{j,m}[n]\}_{L<j\le J,0\le m<2^{-j}}, \{\phi_{J,m}[n]\}_{0\le m<2^{-J}}\right]$. A thresholding estimator in this wavelet basis can be written as

$$\tilde{F} = \sum_{j=L+1}^{J}\sum_{m=0}^{2^{-j}} \rho_\lambda\left(\langle g,\psi_{j,m}\rangle\right)\psi_{j,m} + \sum_{m=0}^{2^{-J}} \rho_\lambda\left(\langle g,\phi_{J,m}\rangle\right)\phi_{J,m}\,.$$

Noise Variance Estimation

The signal f of size N has $N/2$ wavelet coefficients $\{\langle f,\psi_{l,m}\rangle\}_{0\le m<N/2}$ at the finest scale $2^l = 2N^{-1}$. The coefficients $|\langle f,\psi_{l,m}\rangle|$ is small if f is smooth over the support of $\psi_{l,m}$, in which case $\langle f,\psi_{l,m}\rangle \approx \langle \sigma z,\psi_{l,m}\rangle$. In contrast, $|\langle f,\psi_{l,m}\rangle|$ is large if f has sharp transition, and hence produces a number of large coefficients that is small compared to $N/2$. At the finest scale, the signal f thus influences the value of a small portion of large amplitude coefficients $\langle f,\psi_{l,m}\rangle$. All others are approximately equal to $\langle g,\psi_{l,m}\rangle$, which are independent Gaussian random variables

of variance σ^2.

To estimate the variance σ^2 of the noise from the data, we need to suppress the influence of f. When f is piecewise smooth, a robust estimator is calculated from the median of the finest scale wavelet coefficients $\{\langle f, \psi_{l,m} \rangle\}_{0 \le m < N/2}$ [13]. If M is the median of the absolute value of P independent Gaussian random variables of zero-mean and variance σ_o^2, then one can show that $E\{M\} \approx 0.6745\sigma_o$. The variance σ^2 of the noise is estimated from the median M_f of $\{\langle f, \psi_{l,m} \rangle\}_{0 \le m < N/2}$ by neglecting the influence of f,

$$\tilde{\sigma} = \frac{M_f}{0.6745}. \tag{4.8}$$

Indeed f is only responsible for few large amplitude coefficients, and these have little impact on M_f.

Hard or Soft Thresholding

The wavelet thresholding algorithm is shown as the following.

Wavelet thresholding denoising algorithm

1. *Apply J steps of the cascade discrete wavelet transform to get the $N - N/2^J$ wavelet coefficients and $N/2^J$ scaling coefficients corresponding to $g[n]$.*
2. *Choose a threshold $\lambda = \sigma\sqrt{2\log N}$ and apply thresholding (soft or hard) to the wavelet coefficients and leave the scaling coefficients alone.*

3. *Invert the cascade discrete wavelet transform to get the denoised signal.*

Note that the soft thresholding guarantees with a high probability that

$$\left|\left\langle \tilde{F}, \psi_{l,m} \right\rangle\right| = \left|\rho_\lambda\left(\left\langle g, \psi_{l,m} \right\rangle\right)\right| \le \left|\left\langle f, \psi_{l,m} \right\rangle\right|.$$

The estimator $\tilde{F}$ is at least as regular as f because its wavelet coefficients have smaller amplitudes. This is not true for the hard thresholding estimator, which leaves unchanged the coefficients above λ, and which can therefore be larger than those of f because of the additive noise component.

Multiscale SURE Thresholding

Piecewise regular signals have a proportion of large coefficients $\left\langle f, \psi_{l,m} \right\rangle$ that increases when the scale 2^j increases. Indeed, a singularity creates the same number of large coefficients at each scale, whereas the total number of wavelet coefficients increases when the scale decreases. To use this prior information, one can adapt the threshold choice to the scale 2^j. At large scale 2^j the threshold λ_j should be smaller in order to avoid setting too many large amplitude wavelet coefficients of signal to zero, which would increase the risk.

Wavelet thresholding denoising algorithm with SURE thresholds

1. *Apply J steps of the cascade discrete wavelet transform to get the $N - N/2^J$ wavelet coefficients and $N/2^J$ scaling coefficients corresponding to $g[n]$.*

2. *Compute an estimate* $\tilde{\sigma}^2$ *of the noise variance* σ^2 *with the median formula (eqn.4.8) at the finest scale.*

3. *At each scale* 2^j, *calculate a different threshold from the* 2^{-j} *noisy coefficients* $\{\langle f, \psi_{l,m} \rangle\}_{0 \le m < N/2}$ *with the algorithm in section 4.2.2 (eqn.4.4-4.6). (The SURE threshold* λ_j *is calculated by minimizing an estimation (eqn.4.5) of the risk at the scale* 2^{-j}.)

4. *Perform soft thresholding at each scale* 2^{-j} *with the threshold* λ_j.

3. *Invert the cascade discrete wavelet transform to get the denoised signal.*

Note that for hard thresholding, we have no reliable formula with which to estimate the risk and hence compute the adapted threshold with a minimization. We usually simply multiply by 2 the SURE threshold calculated for a soft thresholding.

Translation Invariance

Thresholding noisy wavelet coefficients would create small ripples near discontinuities. Indeed, setting a coefficient $\langle f, \psi_{j,m} \rangle$ subtracts $\langle f, \psi_{j,m} \rangle \psi_{j,m}$ from f, which introduces oscillations whenever $\langle f, \psi_{j,m} \rangle$ is non-negligible. These oscillations are attenuated by a translation invariant estimation (eqn.4.7). Thresholding wavelet coefficients of translated signals and translating back the reconstructed signals yields shifted oscillations created by shifted wavelets that are set to zero. The averaging partially cancels these oscillations, reducing their amplitude.

When computing the translation invariant estimation, instead of shifting the signal, one can shift the wavelets in the opposite direction,

$$\left\langle f[n-p], \psi_{j,m}[n] \right\rangle = \left\langle f[n], \psi_{j,m}[n+p] \right\rangle = \left\langle f[n], \psi_j[n - N2^j m + p] \right\rangle .$$

If f and all wavelets ψ_j are N periodic then all these products are provided by the dyadic wavelet transform defined in section 2.2.3,

$$Wf[2^j, p] = \left\langle f[n], \psi_j[n-p] \right\rangle \qquad \text{for } 0 \le p < N .$$

The "algorithme à trous" of section 2.2.3 computes these $N \log_2 N$ coefficients for $L < j \le 0$ with $O(N \log_2 N)$ operations. One can verify that the translation invariant wavelet estimator (eqn.4.7) can be calculated by thresholding the dyadic wavelet coefficients $\left\langle g[n], \psi_j[n-p] \right\rangle$ and by reconstructing a signal with the inverse dyadic wavelet transform.

4.2.4 Multiwavelet thresholding

Let us first construct a one-dimensional compactly supported multiwavelet transform with multiplicity two. We begin with two vector functions, the orthonormal multiscaling and multiwavelet vector functions which satisfy the following refinement equations.

$$\Phi(t) = 2\sum_{k=0}^{L-1} H_k \Phi(2t-k), \tag{4.8.1}$$

$$\Psi(t) = 2\sum_{k=0}^{L-1} G_k \Phi(2t-k), \tag{4.8.2}$$

where the two matrices $H(\omega) = \sum_{k \in Z} H_k e^{-ik\omega}$ and $G(\omega) = \sum_{k \in Z} G_k e^{-ik\omega}$ are

trigonometric polynomials satisfying

$$H(\omega)H^*(\omega)+H(\omega+\pi)H^*(\omega+\pi)=I_2, \tag{4.9.1}$$

$$H(\omega)G^*(\omega)+H(\omega+\pi)G^*(\omega+\pi)=0_2, \tag{4.9.2}$$

$$G(\omega)G^*(\omega)+G(\omega+\pi)G^*(\omega+\pi)=I_2, \tag{4.9.3}$$

with real coefficients where H^* denotes the complex conjugate transpose of H.

From eqn.4.9, $\{H_k\}_{0\le k\le L-1}$ and $\{G_k\}_{0\le k\le L-1}$ are $M\times M$ filter matrices form an orthogonal FIR multifilter bank and so they generate a compactly supported scaling and multiwavelet vector functions, $\Phi=(\phi_1,\phi_2)^T$ and $\Psi=(\psi_1,\psi_2)^T$, respectively. Let V_j be the multiresolution analysis space generated by Φ, that is, $V_j=span\left\{2^{j/2}\phi_l(2^jt-k)\right\}_{1\le l\le 2,\,k\in Z}$. As $V_{j+1}=V_j\oplus W_j$, W_j is the orthogonal complement of V_j in V_{j+1}, we also have $W_j=span\left\{2^{j/2}\psi_l(2^jt-k)\right\}_{1\le l\le 2,\,k\in Z}$.

Definition 4.1 ***Continuous Multiwavelet Transform***

For any continuous function $f(t)$ *in* V_0, *it can be expanded as* $f(t)=\sum_n\left(c_{1,n}^{(0)}\phi_1(t-n)+c_{2,n}^{(0)}\phi_2(t-n)\right)$. *The function* f *is completely determined by the sequence* $\left\{(c_{1,n}^{(0)},c_{2,n}^{(0)})^T\right\}$. *Let* J *be the lowest number of decomposition level which is negative. Since* $V_0=W_{-1}\oplus V_{-1}=\cdots=W_{-1}\oplus\cdots\oplus W_J\oplus V_J$ *and* $2^{j/2}\phi_l(2^j-k)$ *and* $2^{j/2}\psi_l(2^j-k)$ *are orthonormal to each other,* f *can also be expanded as,*

$$f(t)=\sum_n\left(\sum_{l=1}^{2}c_{l,n}^{(-1)}\frac{1}{\sqrt{2}}\phi_l\left(\frac{t}{2}-n\right)+\sum_{l=1}^{2}d_{l,n}^{(-1)}\frac{1}{\sqrt{2}}\psi_l\left(\frac{t}{2}-n\right)\right)$$

$$=\sum_n\left(\sum_{l=1}^{2}c_{l,n}^{(J)}2^{J/2}\phi_l(2^Jt-n)+\sum_{j=J}^{-1}\sum_{l=1}^{2}d_{l,n}^{(j)}2^{j/2}\psi_l(2^jt-n)\right).$$

where $c_{l,k}^{(j)}$ *and* $d_{l,k}^{(j)} \in \Re$ *are given by*

$$c_{l,k}^{(j)} = \int f(t) 2^{j/2} \phi_l\left(2^j t - k\right) dt,$$

$$d_{l,k}^{(j)} = \int f(t) 2^{j/2} \psi_l\left(2^j t - k\right) dt.$$

Univariate VS Bivariate Thresholding

Strela *et al.* applied the single wavelet thresholding method (*univariate*) to a GHM multiwavelet decomposition to produce some encouraging signal denoising results [49]. Even though univariate thresholding does work in multiwavelet denoising, further noise reduction is possible. This is because multiwavelet transform normally produces correlated vector coefficients. If there is a signal component present at a particular time-frequency location, then we would also expect the other element in that vector coefficient contains signal component. Any prefilter except the identity prefilter gives these correlated coefficients, but the identity prefilter gives poor denoising results with the thresholding approach. Therefore, Downie and Silverman proposed to use a thresholding method that will treat the multiwavelet vector coefficient as a whole entity [18]. Accordingly, the *multivariate* (for multiplicity 2, *bivariate*, that is, using two scaling and wavelet functions) thresholding method accounts for the noise and signal components within the whole vector. They also showed that the repeated signal prefilter is very good for denoising applications [18].

The theory behind bivariate thresholding is briefly reviewed as the following.

Suppose we apply the discrete multiwavelet transform in definition 4.1 with an appropriate prefilter to a noisy signal, then we obtain two streams of vector coefficients in the form $D_{j,k} = D^*_{j,k} + E_{j,k}$, where $D^*_{j,k}$ is the vector signal coefficient and $E_{j,k}$ has a multivariate normal distribution $N_L(0, C_j)$. The matrix C_j is the covariance matrix for the error term that depends on the resolution scale j. Using the standard transform $\theta^2_{j,k} = [D_{j,k}]^T C_j^T D_{j,k}$, we obtain a positive scalar value that, in the absence of any signal component, will have a X^2_L distribution. The threshold rule is based on these values. To find the universal threshold, we choose a sequence such that $\lim_{n\to\infty} \Pr(M_n < \lambda_n)$ is strictly between 0 and 1, where M_n is the maximum i.i.d. X^2_M random variables. An appropriate threshold sequence is $\lambda_n = 2\log n + (M-2)\log(\log n)$. When $M = 2$, the universal threshold for multiple wavelets simplifies to $\lambda_n = 2\log n$. Unlike the single wavelet thresholding, the variance term σ does not appear in the universal threshold formula. Then the hard thresholding rule in bivariate thresholding can be written as

$$\hat{D}_{j,k} = \begin{cases} D_{j,k} & \text{if } \theta_{j,k} \geq \lambda_n \\ 0 & \text{otherwise} \end{cases}.$$

The bivariate soft thresholding can be formulated as

$$\hat{D}_{j,k} = \begin{cases} D_{j,k}\left(1 - \dfrac{\lambda_n}{\theta_{j,k}}\right) & \text{if } \theta_{j,k} \geq \lambda_n \\ 0 & \text{otherwise} \end{cases}.$$

To compute $\theta_{j,k}$, we need C_j, which can be obtained from $C_j = Var(D_{j,k})$. The

covariance structure of the wavelet coefficients at each level can be obtained explicitly.

Translation Invariant VS non-Translation Invariant Multiwavelets

It was concluded by Bui and Chen in [2] that translation invariant multiwavelet denoising is generally better than non-translation invariant multiwavelet denoising, no matter what thresholding method is used. However, it was found that translation invariant multiwavelet denoising does not always outperform translation invariant single wavelet denoising. They also verified that signal denoising using bivariate thresholding is better than using univariate thresholding.

4.3 The proposed algorithm

In [56], vector-valued multiresolution analysis (VMRA) for vector-valued signals is introduced. The vector-valued scaling and wavelet functions are also defined. They can be generated from some lowpass and bandpass filters in matrix form, which is called matrix quadrature mirror filter (MQMF). Multiwavelet transform includes the prefiltering and postfiltering process, which vectorize a scalar input signal for VMRA. The concepts of orthogonality and orthonormal bases for multiwavelet transform are similar to the traditional scalar wavelet transform in Hilbert spaces. The matrix dilation equations are also similar to those in the traditional scalar wavelet transform. As the component functions in vector-valued wavelets can form multiwavelets, certain linear combinations of scalar-valued wavelets can yield multiwavelets.

Suppose that the multiwavelet ψ_l is a complex-valued function on $\Re$ that has n_l vanishing moments and is n_l times continuously differentiable with derivatives that also have a fast decay. This means that for any $0 \le k \le n$ and $m \in \mathrm{N}$ there exists continuous function $C_{l,q}$ such that

$$\forall t \in \Re, \qquad |\psi_l(t)| \le \frac{C_{l,q}}{1+|t|^q}. \tag{4.10}$$

Now multiwavelet functions ψ_1 and ψ_2 have n_1 and n_2 vanishing moments respectively. Multiwavelet Ψ has $n = \min(n_1, n_2)$ vanishing moments is

orthogonal to polynomials of degree $\min(n_1-1, n_2-1)$. When we estimate the Lipschitz exponents of f, we also ignore the wavelet transform of the Taylor polynomial p_v in section 2.2.4 if we use a multiwavelet Ψ that has $n > \lfloor \alpha \rfloor$ vanishing moments. It is because if $n > \lfloor \alpha \rfloor$, polynomial p_v has degree at most $n-1$ and its multiwavelet transform also vanishes.

$$\text{For } l = 1,2, \qquad \int_{-\infty}^{\infty} p_v(t) 2^{j/2} \psi_l(2^j t - k)\, dt = 0. \tag{4.11}$$

The number of high amplitude wavelet coefficients created by a brutal transition like an edge is proportional to the width of the supports of the filters. For a more accurate localization of singularities, the number of large amplitude wavelet coefficients produced should be as small as possible. So the supports of the filters should be as short as possible. Multifilters are possible to have shorter supports and more vanishing moments. Moreover, the more the vanishing moments, the smaller the coefficients can be produced at fine scales. Therefore, the multiwavelet coefficients that belong to the noise component should be more easily distinguished at fine scales. Apart from this, computational complexity can be greatly reduced by multiwavelet transform. The relation between uniform and pointwise Lipschitz regularity and wavelet transform has already been proposed and proved in [30,72,74,75,76,80]. Regularity is measured by looking at the asymptotic decays of wavelet transform coefficients instead of Fourier transform coefficients. This is because wavelets can

simultaneously localize signals in time and frequency domains. The property of localization in time enables us to estimate local regularity, whereas the localization in frequency enables the measurement. Multiwavelets are generated by several scaling functions. Preprocessing is necessary when applying discrete multiwavelet filter bank to scalar signals. The combined filter responses may not provide localization in frequency domain. However, it can be shown that with an appropriate designed prefilter and postfilter, one can estimate the Lipschitz regularity from multiwavelet coefficients. In theorem 4.1, we relate the uniform Lipschitz regularity of a signal f on an interval $[a,b]$ to the amplitude of its multiwavelet transform [31].

Definition 4.2 ***Discrete Multiwavelet Transform***

Let us denote $c_n^{(j)} = \left(c_{1,n}^{(j)}, c_{2,n}^{(j)}\right)^T$ *and* $d_n^{(j)} = \left(d_{1,n}^{(j)}, d_{2,n}^{(j)}\right)^T$, *we have the decomposition algorithm as,*

$$\sum_n \left(c_n^{(j)}\right)^T 2^{j/2}\,\Phi\left(2^j t - n\right) = \sum_n \left(\left(c_n^{(j-1)}\right)^T 2^{j-1/2}\,\Phi\left(2^{j-1} t - n\right) + \left(d_n^{(j-1)}\right)^T 2^{j-1/2}\,\Psi\left(2^{j-1} t - n\right) \right),$$

with the following dilation equations

$$c_k^{(j-1)} = \sqrt{2} \sum_n H_{n-2k} c_n^{(j)}, \tag{4.12.1}$$

and

$$d_k^{(j-1)} = \sqrt{2} \sum_n G_{n-2k} c_n^{(j)}. \tag{4.12.2}$$

At the synthesis bank, we have the reconstruction algorithm as

$$c_k^{(j)} = \sqrt{2} \sum_n \left(H_{k-2n}^T c_n^{(j-1)} + G_{k-2n}^T d_n^{(j-1)} \right). \tag{4.13}$$

These relations enable us to construct a multi-input multi-output filter bank with

multiplicity two.

Theorem 4.1

If $f \in L^2(\Re, C)$ *is uniformly Lipschitz* α *over* $[a,b]$, *then there exists a finite constant* A *such that* $\forall (n,2^{-j}) \in [a,b] \times \Re^+$,

$$\left|d_{l,n}^{(j)}\right| \le A\left(2^{-j}\right)^{\alpha+1/2}, \tag{4.14}$$

for $l = 1,2$, *where* $d_{l,n}^{(j)}$ *is defined in definition 4.1. Conversely, suppose that* f *is bounded and for* $l = 1,2$, $d_{l,n}^{(j)}$ *satisfies eqn.4.14 for an* α *that is not an integer. Then* f *is uniformly Lipschitz* α *on* $[a,b]$.

The proof is shown in Appendix II.

To relate the Lipschitz regularity of f on an interval $[a,b]$ to the magnitude sum of its discrete multiwavelet transform over the COI, we need the pointwise Lipschitz regularity condition [31].

Theorem 4.2

If $f \in L^2(\Re, C)$ *is pointwise Lipschitz* α *at* v, *then there exists a finite constant* A *such that*

$$\forall (n,2^{-j}) \in \Re \times \Re^+, \qquad \left|d_{l,n}^{(j)}\right| \le A\left(2^{-j}\right)^{\alpha+1/2}\left(1+\left|n-2^j v\right|^{\alpha}\right). \tag{4.15}$$

Conversely, if α *is not an integer and there exists a finite constant* A *and* $\alpha' < \alpha$ *such that*

$$\forall (n,2^{-j}) \in \Re \times \Re^+, \qquad \left|d_{l,n}^{(j)}\right| \le A\left(2^{-j}\right)^{\alpha+1/2}\left(1+\left|n-2^j v\right|^{\alpha'}\right), \tag{4.16}$$

Then f *is Lipschitz* α *at* v.

The proof is shown in Appendix III.

By theorem 4.1, we can characterize regularity from the evolution of the magnitudes of the multiwavelet coefficients across scales. Different extents of regularity, which are represented by the Lipschitz exponents, can be estimated from the interscale evolution of the magnitudes of the multiwavelet coefficients. By using the interscale ratio as the criteria to perform the thresholding of the multiwavelet coefficients, we can reject those coefficients corresponding to the noise component. We compute the interscale ratio threshold as the following,

$$R_{l,n}^{(j)} = \frac{d_{l,n}^{(j)}}{d_{l,n}^{(j+1)}} \leq \frac{A\left(2^{-j}\right)^{\alpha}}{A\left(2^{-(j+1)}\right)^{\alpha}} = 2^{\alpha}. \tag{4.17}$$

For multiwavelets case, we estimate the Lipschitz regularity from the multiwavelet transform magnitude sum over the COI and perform denoising by thresholding the multiwavelet coefficients according to their interscale ratios and difference of the multiwavelet transform magnitude sums. Let the support of ψ_l be $[-K_l, K_l]$ for $l = 1,2$. The COI of v in the scale-space plane, $COI_{l,v}^{(j)}$, is the set of index pairs $\left(n, 2^{-j}\right)$ such that n is included in the support of ψ_l. Since the support of ψ_l is equal to $\left[(v-K_l)2^{-j}, (v+K_l)2^{-j}\right]$, the COI of v is defined by $\left|n - 2^{-j}v\right| < K_l 2^{-j}$, that is $COI_{l,v}^{(j)} = \left\{\left(n, 2^{-j}\right) : \left|n - 2^{-j}v\right| \leq K_l 2^{-j}\right\}$. The magnitude sums are computed as

$$N_{l,v}^{(j)} f = \sum_{(l,n)\in COI_{l,v}^{(j)}} \left| d_{l,n}^{(j)} \right|. \tag{4.18}$$

The COI of an abscissa v consists of the scale-space points $\left(n,2^{-j}\right)$ for which the support of ψ_l intersects $t=v$. If n is inside the COI of v then $d_{l,n}^{(j)}$ depends on the value of f in the neighborhood of v. Since $|n-v|/2^{-j} < K_l$, the pointwise Lipschitz condition in theorem 4.2 can be written $\left|d_{l,n}^{(j)}\right| \le A'\left(2^{-j}\right)^{\alpha+1/2}$, which is identical to the uniform Lipschitz condition in theorem 4.1.

The denoising algorithm is simple and easy to be implemented. The procedures are listed as the following.

Multiwavelet denoising algorithm using singularity detection (basic)

1. *Perform the discrete multiwavelet transform (five levels of decomposition) after prefiltering the noisy test signal (eqn. 4.12 and 4.13).*
2. *Determine the COI for all scales and compute the magnitude sum of the transformed coefficients over the COI for all positions.*
3. *Compute the interscale ratio of the magnitude sum at each position. Retain those multiwavelet coefficients with values higher than 2.*
4. *Reconstruct the signal from the selected multiwavelet coefficients using the inverse discrete multiwavelet transform and postfiltering.*

Remarks

1. *As it is observed by the authors in* [32] *that WTMS of some small irregular*

signals with Lipschitz exponents between -1 *and* 0 *will also increase as the scale increases. This enables them to falsely fulfill the interscale ratio thresholding criterion. So the interscale difference thresholding criterion was also employed here.*

2. *Two approaches of retaining coefficients were experimented. They are retaining the coefficients within the COI and retaining only the maximal coefficients (coefficients at the position of the maxima of transformed impulse) inside the COI.*

3. *The suitable threshold values for the interscale ratio is* 2 *and the interscale difference is 0.1 for retaining the coefficients within the COI. While they are respectively* $\sqrt{2}$ *and 0.01 for retaining the maximal coefficients inside the COI.*

4. *Both of the approaches use a simple joint selection, that is, consider the magnitude sum of the square sum of the magnitudes of the two coefficient components.*

4.4 Results

The multiwavelet transform of the noisy signal is shown in figure 4.2. The selected transform coefficients are shown in figure 4.3. We can clearly see that those coefficients belonging to the noise components are distinguished and rejected. The reconstructed denoised signal is shown in figure 4.4(b) and figure 4.5(b). As a preliminary work, we experimented the above denoising algorithm to a length-256 1-d noisy signal using the Xia's prefilter [59] and the GHM multifilters [21] (appendix IV). The original signal contains various features such as steps, regular slope, steep slopes and irregular structures. White Gaussian noise with noise variance equal to 0.1 is added to the original signal to produce a noisy signal with MSE equal to 0.0085 (figure 4.4(a)). 30.63% reduction in MSE was obtained (the MSE decreases from 0.007387 to 0.005124) by the previous denoising algorithm using the single wavelet singularity detection [88]. In our experiment, 34.1% reduction in MSE can be obtained (the MSE decreases from 0.0085 to 0.0056) by the denoising algorithm using the multiwavelet singularity detection [87].

We also experimented the same original signal corrupted by white Gaussian noise with noise variance equal to 0.05 and impulsive noise to produce a noisy signal with a MSE equal to 0.0343 (figure 4.5(a)). We used the Plonka's prefilter [43] and the GHM multifilters [21] (appendix IV) for the multiwavelet decomposition. 77.96% reduction

in MSE was obtained (the MSE decreases from 0.036909 to 0.008133) by the previous denoising algorithm using the single wavelet singularity detection in [32]. In our experiment, 81.04% reduction in MSE can be obtained (the MSE decreases from 0.0343 to 0.0065) by the denoising algorithm using the multiwavelet singularity detection [87]. The reconstructed denoised signal is shown in figure 4.5(b).

We found that retaining those coefficients within the COI is effective for denoising the white Gaussian noise, while retaining those maxima coefficients (remark 2) in the COI is more effective for denoising the white Gaussian noise and impulsive noise. Moreover, it was found that performing selection in a simple joint manner (remark 4) is better than in an individual manner.

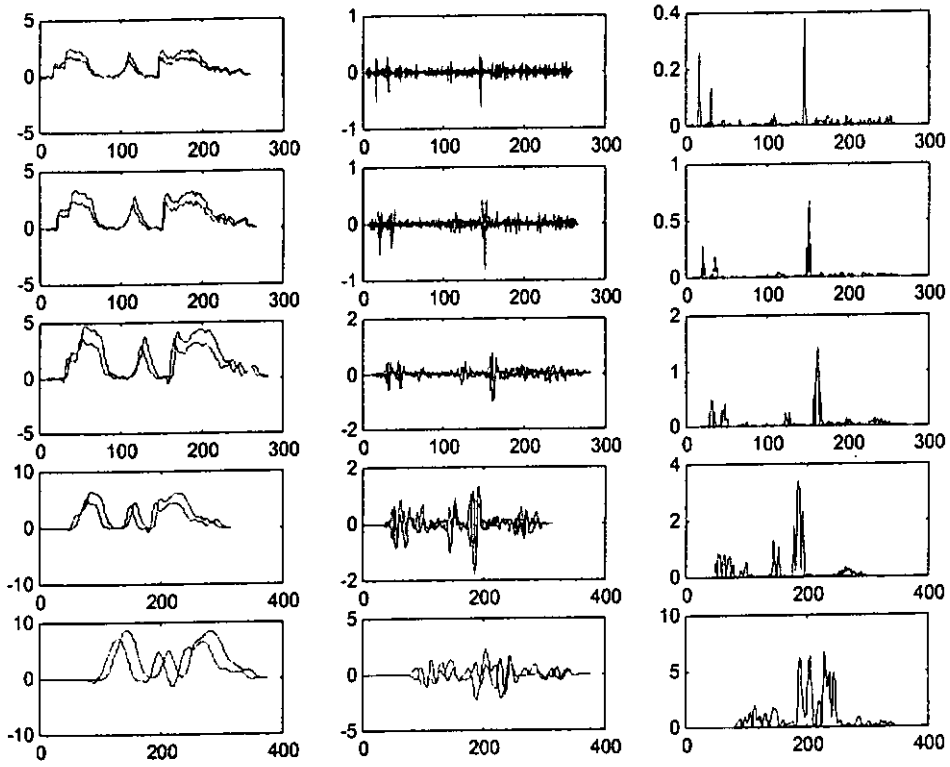


Figure 4.2 The multiwavelet transform of the noisy signal. The 1st column and the 2nd column are the two transform components. The 3rd column is the modulus of these two components. Down the rows is the decreased scales of the multiwavelet transform.

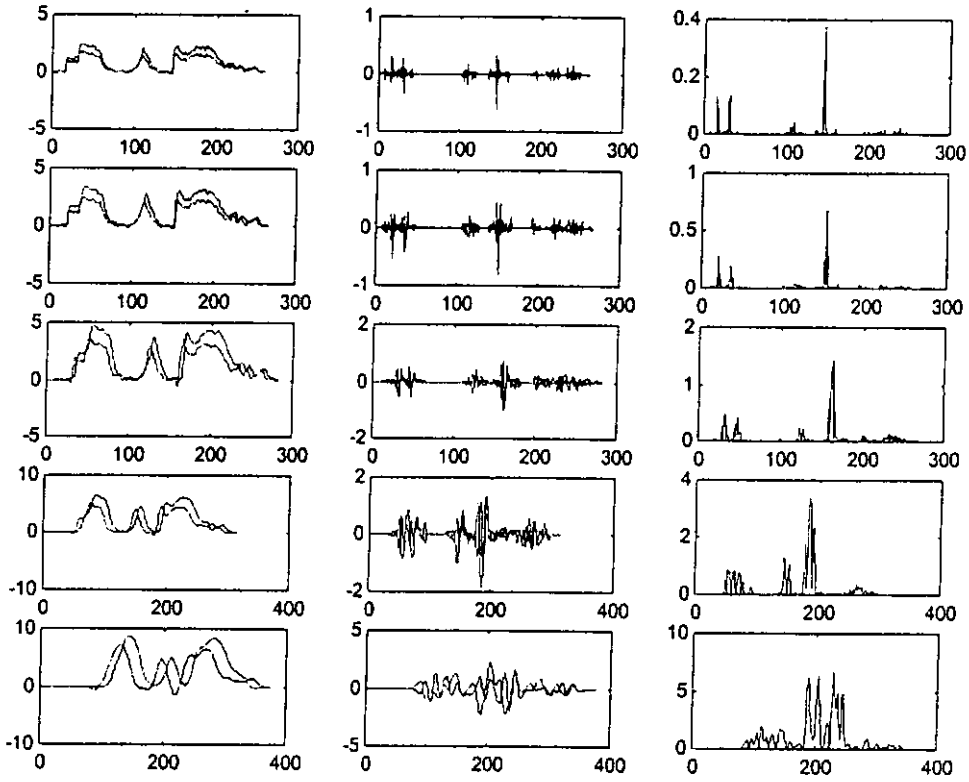
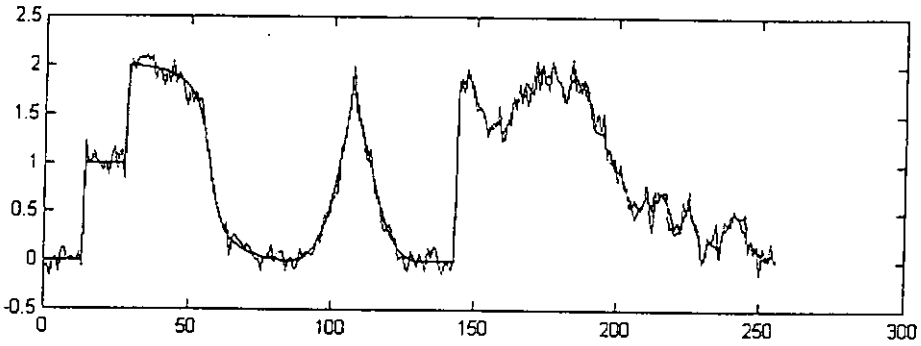
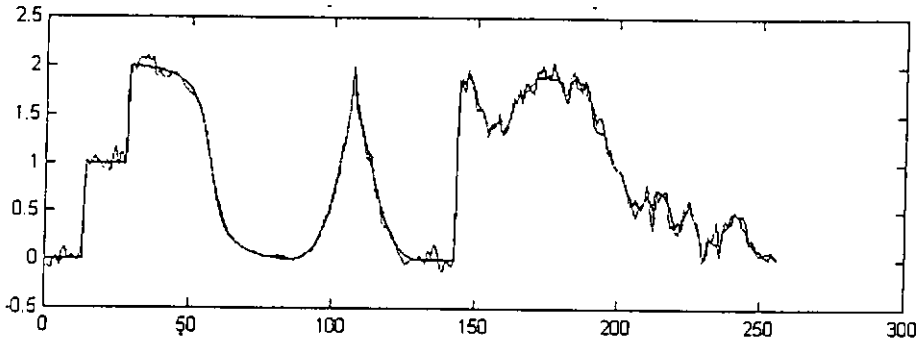


Figure 4.3 The multiwavelet transform of the denoised signal. The 1st column and the 2nd column are the two transform components. The 3rd column is the modulus of these two components. Down the rows is the decreased scales of the multiwavelet transform.

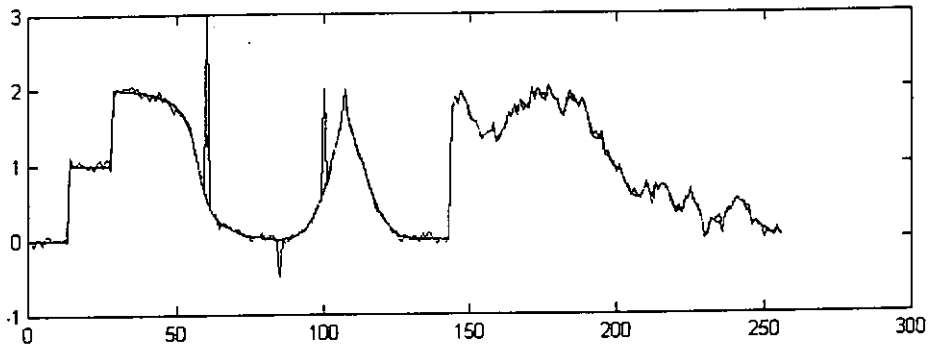


(a) MSE=0.0085

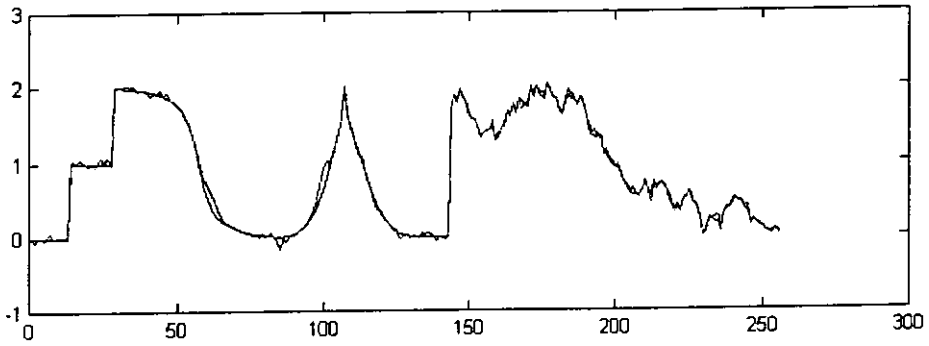


(b) MSE=0.0056

Figure 4.4 (a) The noisy signal corrupted by white Gaussian noise, (b) the denoised signal by the denoising algorithm using multiwavelet singularity detection.



(a) MSE=0.0343



(b) MSE=0.0065

Figure 4.5 (a) The noisy signal corrupted by white Gaussian noise and impulsive noise, (b) the denoised signal by the denoising algorithm using multiwavelet singularity detection.

4.5 Summary

In this chapter, we extended the singularity detection from the single wavelet transform to the multiwavelet transform. We first reviewed the existing wavelet and multiwavelet denoising techniques. We then developed the multiwavelet singularity detection and implemented a signal denoising algorithm.

From the preliminary denoising results of 1-d white Gaussian noise and impulsive noise corrupted signals, we verified the advantageous properties of the multiwavelet transform over the single wavelet transform for singularity detection. First of all, the shorter supports of the multifilters contribute to the improvement of the singularity detection, as the width of the COI becomes narrower and so the overlappings of the non-isolated transformed singularities become less. The vanishing moments of the multifilters is also high enough to provide good signal approximation, so that the wavelet coefficients produced at fine scales become smaller and a higher extent of noise attenuation upon thresholding can be obtained. Although multiwavelet transform requires more memory than single wavelet transform due to parallel vector processing, and it has a higher computational complexity [96], we still conclude that signal denoising using the multiwavelet singularity detection is better than signal denoising using the single wavelet singularity detection.

Chapter 5

Conclusion

5.1 General conclusion

In this work, we presented the application of the wavelet singularity detection to a regularity scalable image coding algorithm. We further extended the singularity detection to the multiwavelet transform. A signal denoising algorithm was then implemented to verify this result. In this section, we draw some conclusions in these aspects.

First of all, we reviewed the history and development of the wavelet transform in chapter 2. We started from Fourier transform to the time-frequency wedding and the windowed Fourier transform, which are not able to provide sufficient time-frequency localization for transient signals. Then we reviewed the introduction of the wavelet transform, and its relation with the discrete implementation by filter banks. We highlighted the difference between the continuous time wavelet theory and the discrete filter banks, which become equivalent through the elaboration of orthogonal wavelets and multiresolution approximations. We then described the important features of the wavelet transform. For the realization of the wavelet transform, fast

dyadic transform and the *algorithme à trous* were first introduced. In the discrete implementation, we concerned the filter banks with perfect reconstruction. It is because the signals involved in the denoising algorithms and the proposed scalable coding algorithm using singularity detection are decomposed, processed and simply reconstructed by the inverse transform. We also went through different boundary filtering techniques for finite signals. Next, we reviewed the wavelet singularity detection. Lipschitz regularity of continuous time signal was defined and the estimation of Lipschitz exponents from the evolution of the oriented wavelet transform were presented. We reviewed two approaches for the estimation of the Lipschitz regularity – the wavelet transform modulus maxima and the wavelet transform modulus sum. We listed out the applications of these theories in edge detection and signal and image denoising.

Since we extended the singularity detection from the single wavelet transform to the multiwavelet transform, we also reviewed the background of the multiwavelet transform in chapter 2. We outlined the important and specific features that can only be simultaneously obtained from the multiscaling functions. We reviewed the concept of the prefilter design and the vector filter bank implementation for the realization of the multiwavelet transform.

Chapter 3 is the core of this dissertation. We designed and implemented a regularity scalable image coding algorithm which is adaptive to image signal. It can avoid the disadvantages occurred in the existing feature-based scalable image coding algorithms. We first reviewed the resolution and SNR scalable coding algorithms, the edge-enhanced coding algorithm, the edge-oriented scalable coding algorithm and the surface-oriented scalable coding algorithm. Since the existing WTMS approach cannot directly be applied to our proposed algorithm, where the separable wavelet transform with decimation is used to decompose image signal in a wavelet image code, we developed the singularity detection by the magnitude sum over the decimated COI. A wavelet image decomposition provides a hierarchical data structure, and this selection utilizes the tree structured data organization. With partial ordering of the wavelet coefficient magnitudes, the entropy of the selected coefficients at the first few regularity levels are lowered and the compression rate is increased. The proposed algorithm and the results were presented. Compared to the performance of the wavelet image coder with resolution scalability only, we found that 44.25% reduction in bit rate and only 2.86% reduction in PSNR, can be obtained by the proposed algorithm when it is combined with resolution scalability. Here the test image "Lena" is encoded with quantization step size 6 at resolution scale 2. Zooming and observing the difference of the decoded images, we also found that the quality of the strings on the

hat are better as a result of bits being assigned to code that level of regularity first. While it is not the case for the smooth shoulder area, where the degradation is severe but it is not noticeable. The quality of the stripe pattern of the hat is lying between them. Compared to the other feature-based scalable wavelet image coding algorithm, improved visual quality of the reconstructed images at different bit rates can be obtained.

As a further development of the application of wavelet singularity detection, we extended the singularity detection to the multiwavelet transform. In chapter 4, we presented the preliminary results. We first reviewed the signal and image denoising using thresholding techniques. This includes the thresholding estimation and refinement and the previous works of wavelet thresholding. The univariate thresholding was presented as an initial work of multiwavelet denoising and later the bivariate thresholding, which utilizes the multiplicity of multiwavelet transform, was presented. We also present the importance of translation invariance in the single wavelet and multiwavelet denoising. To demonstrate the advantages of the multiwavelet transform, we implemented a denoising algorithm based on multiwavelet singularity detection. We showed the Lipschitz regularity condition of the discrete multiwavelet transform. Based on this result we implemented a signal denoising algorithm. Preliminary results showed that the denoising algorithm using

multiwavelet singularity detection (34.1% reduction in MSE) is better than the denoising algorithm using single wavelet singularity detection (30.63% reduction in MSE) for a white Gaussian noise corrupted signal. For a white Gaussian and impulsive noise corrupted signal, 81.04% reduction in MSE can be obtained by the denoising algorithm using multiwavelet singularity detection, which is also better than the denoising algorithm using single wavelet singularity detection (77.96% reduction in MSE).

We accounted for the above results as being due to the shorter supports and high vanishing moments that the multifilters can achieve. Intuitively, the shorter filter supports means the narrower COI and better frequency localization that the magnitude sum can have for measuring the regularity using their interscale ratios. It is because the width of the COI becomes smaller and so the overlappings of the non-isolated transformed singularities become less. The vanishing moments of the multifilters is also high enough to have good signal approximation, so that the wavelet coefficients produced at fine scales become smaller and a higher extent of noise attenuation upon thresholding can be obtained. Moreover, the correlated vector transform coefficients produced by the preprocessing also provide more information on the regularity of the signal.

5.2 Future developments

The results obtained in this work are in fact small contributions to the development of wavelet singularity detection for signal and image processing. Yet our works provides plausible solutions to the problems mentioned above and introduces new ideas for further research. In this section, we discuss some of the possible future developments.

There are some tasks that are going to be completed in short term. To make the work more complete, we need the derivation of the Lipschitz condition for the multiwavelet transform magnitude sum, for the joint coefficient components, such as the square sum of the magnitudes of the two components, and the individual coefficient components. For the denoising algorithm, we can further investigate the individual and joint selection of the coefficients. For the joint selection, we need to define good criteria to combine the components of the coefficient vector. Square sum is the simplest but representative combination. We can also trace the evolution of the argument between these components. The criteria can then be applied to the selection of coefficients by retaining those coefficients within the COI or those coefficients located at the maxima of the transformed impulse inside the COI. With some experiments on these different criteria we can then develop the best algorithm for image denoising. To embed the denoising algorithm in a multiwavelet image coder [90], we have to investigate the singularity detection and coefficient selection

algorithm for the decimated case.

In long term we can explore the following areas. Since prefiltering produces correlated coefficients which contain information of the regularity of an input signal, we can investigate how to capture this information from the combined filtering and characterize it. This can be achieved by suitably designing a set of prefilter and postfilter with this eligibility. We can also experiment the algorithm developed above with different existing prefilters and multifilters. Actually the multiwavelet filter and prefilter design is a broad and deep topic. We can study and develop a new set of prefilter and multifilters which is specific to our purposes.

Appendix I

Proof of the necessary condition (eqn.3.2) in theorem 3.1

Since f *is Lipschitz* α *in* $U(v)$, f *is also Lipschitz* α *at* v, $\forall v \in U(v)$, *by eqn.2.37, there exists a polynomial* p_v *of degree* $n > \lfloor \alpha \rfloor$ *and* $K > 0$ *such that*

$$|f(x_1,x_2) - p_v(x_1,x_2)| \le K\left(|x_1 - v_1|^2 + |x_2 - v_2|^2\right)^{\frac{\alpha}{2}}. \qquad \text{(A1.1)}$$

$$\begin{aligned} |W_j^l f(u)| &= \left| \int\int_{-\infty}^{\infty} f(x_1,x_2)\psi_{j,u}^l(x_1,x_2)\,dx_1\,dx_2 \right| \\ &= \left| \int\int_{-\infty}^{\infty} (f(x_1,x_2) - p_v(x_1,x_2))\psi_{j,u}^l(x_1,x_2)\,dx_1\,dx_2 \right|, \quad \textit{by eqn.2.33,} \\ &\le \int\int_{-\infty}^{\infty} |f(x_1,x_2) - p_v(x_1,x_2)|\,|\psi_{j,u}^l(x_1,x_2)|\,dx_1\,dx_2 \\ &\le \int\int_{-\infty}^{\infty} K\left(|x_1 - v_1|^2 + |x_2 - v_2|^2\right)^{\frac{\alpha}{2}} |\psi_{j,u}^l(x_1,x_2)|\,dx_1\,dx_2, \ \textit{by eqn.A1.1.} \end{aligned}$$

By Cauchy Schwarz Inequality, $|a+b|^r \le 2^r\left(|a|^r + |b|^r\right)$,

$$|W_j^l f(u)| \le A \int\int_{-\infty}^{\infty} \left(|x_1 - v_1|^\alpha + |x_2 - v_2|^\alpha\right) |\psi_{j,u}^l(x_1,x_2)|\,dx_1\,dx_2.$$

With the change of variables, $x_1' = 2^{-j}x_1 - u_1$ *and* $x_2' = 2^{-j}x_2 - u_2$, *we have*

$$\begin{aligned} |W_j^l f(u)| \le A(2^j)^\alpha \Bigg(& \int\int_{-\infty}^{\infty} |x_1'|^\alpha \, |\psi^l(x_1',x_2')|\,dx_1'\,dx_2' + \int\int_{-\infty}^{\infty} |x_2'|^\alpha \, |\psi^l(x_1',x_2')|\,dx_1'\,dx_2' \\ & + |u_1 - 2^{-j}v_1|^\alpha \int\int_{-\infty}^{\infty} |\psi^l(x_1',x_2')|\,dx_1'\,dx_2' + |u_2 - 2^{-j}v_2|^\alpha \int\int_{-\infty}^{\infty} |\psi^l(x_1',x_2')|\,dx_1'\,dx_2' \Bigg) \end{aligned}$$

Since we usually just consider the vanishing moment of the mother wavelet $\psi(x)$, *so* $\psi^l(x)$ *has* n *vanishing moments for all* l, *and is* C^n *with derivatives that have a fast decay. So by eqn.2.34, we have*

$$|W_j^l f(u)| \le A(2^j)^\alpha \left(2 + |u_1 - 2^{-j}v_1| + |u_2 - 2^{-j}v_2|\right) \qquad \text{(A1.2)}$$

For $u \in U(v)$, *we can choose* $v = (2^{-j}v_1, 2^{-j}v_2) = u$, *so we have* $|W_j^l f(u)| \le A(2^j)^\alpha$.

Proof of the sufficient condition (eqn.3.3) of theorem 3.1

To prove that f *is uniformly Lipschitz* α *in* $U_\delta(v)$, *we must verify that there exists* $K > 0$ *such that for all* $v \in U_\delta(v)$, *we can find a polynomial* p_v *of degree* $\lfloor \alpha \rfloor$ *such that* $\forall x, v \in \Re^2$,

$$|f(x_1, x_2) - f(v_1, v_2)| \le K\left(|x_1 - v_1|^2 + |x_2 - v_2|^2\right)^{\frac{\alpha}{2}}. \qquad \text{(A1.3)}$$

<u>*Case 1*</u> $x \notin U_{\delta/2}(v)$

Then $|x_1 - v_1| \ge \delta/2$ *and* $|x_2 - v_2| \ge \delta/2$.

Since f *is bounded, eqn.A1.3 is verified with a constant* $K > 0$ *that depends on* δ.

<u>*Case 2*</u> $x \in U_{\delta/2}(v)$

Since f *can be decomposed in a Littlewood-Paley type Sum*

$$f(x) = \sum_{j \to -\infty}^{+\infty} \Delta_j(x) \qquad \text{(A1.4)}$$

with $$\Delta_j(x) = \frac{1}{C_{\psi^l}} \int\int_{-\infty}^{+\infty} \int_{2^j}^{2^{j+1}} W_j^l f(u) \frac{1}{s} \psi^l\left(2^{-j}x - u\right) \frac{ds}{s^2} du_1 du_2, \quad \text{(A1.5)}$$

where $C_{\psi^l} = C_\phi C_\psi$ *or* $C_{\psi^l} = C_\psi C_\psi$, $C_\psi = \int_{-\infty}^{+\infty} \frac{|\hat{\psi}(w)|^2}{w} dw < +\infty$ *is a weak admissibility condition by Calderón [3] and* $s = 2^j$.

Let $(D_i)^r = \frac{\partial^r}{\partial x_i^r}$ *be a differential operator and* $D_1^{k-r} D_2^r \Delta_j$ *be the kth derivative of* Δ_j. *To prove that* f *is uniformly Lipschitz* α *in* $U_\delta(v)$, *we shall approximate* f *with a polynomial* $p_v(x) = f(v)$ *that generates the Taylor polynomial,*

$$p_v(x)=\sum_{j\to-\infty}^{+\infty}\sum_{k=0}^{\lfloor\alpha\rfloor}\frac{\left[\left((x_1-v_1)D_1+(x_2-v_2)D_2\right)^k\Delta_j\right](v)}{k!}. \qquad \text{(A1.6)}$$

Since $\psi^1(x)$ *has fast decay, by eqn.2.34 and eqn.A1.2, we have*

$$\left|\Delta_j(x)\right|\le\frac{1}{C_\psi}\int\int_{-\infty}^{+\infty}\int_{2^j}^{2^{j+1}}As^\alpha\left(2+\left|u_1-2^{-j}v_1\right|^{\alpha'}+\left|u_2-2^{-j}v_2\right|^{\alpha'}\right)\frac{C_q}{1+\left(2^{-j}x_1-u_1\right)^q+\left(2^{-j}x_2-u_2\right)^q}\frac{ds}{s^2}du_1du_2$$

$$=K\left(2^j\right)^\alpha\int\int_{-\infty}^{+\infty}\frac{2+\left|u_1-2^{-j}v_1\right|^{\alpha'}+\left|u_2-2^{-j}v_2\right|^{\alpha'}}{1+\left(2^{-j}x_1-u_1\right)^q+\left(2^{-j}x_2-u_2\right)^q}\frac{du_1du_2}{2^j}. \qquad \text{(A1.7)}$$

By the inequality $\left|a-b\right|^r\le2^r\left(\left|a-c\right|^r+\left|c-b\right|^r\right)$, *with* $a=u$, $b=2^{-j}v$ *and* $c=2^{-j}x$, *and with the change of variables,* $u_1'=u_1-2^{-j}x_1$ *and* $u_2'=u_2-2^{-j}x_2$,

$$\left|\Delta_j(x)\right|\le K\left(2^j\right)^\alpha\int\int_{-\infty}^{+\infty}\frac{2+\left|u_1'\right|^{\alpha'}+\left|u_2'\right|^{\alpha'}+\left|2^{-j}(x_1-v_1)\right|^{\alpha'}+\left|2^{-j}(x_2-v_2)\right|^{\alpha'}}{1+\left(u_1'\right)^q+\left(u_2'\right)^q}du_1'du_2'$$

After choosing $m=\alpha'+2$, *we have,*

$$\left|\Delta_j(x)\right|\le K\left(2^j\right)^\alpha\left[2+2^{\alpha'}\left(\left|2^{-j}(x_1-v_1)\right|^{\alpha'}+\left|2^{-j}(x_2-v_2)\right|^{\alpha'}\right)\right]. \qquad \text{(A1.8)}$$

Note that this is also applied to the partial derivatives of $\Delta_j(x)$, $\forall k\le\lfloor\alpha\rfloor+1$,

$$\left|D_1^{k-r}D_2^r\Delta_j(x)\right|\le K\left[\left(2^j\right)^{\alpha-k+r}\left(1+\left|2^{-j}(x_1-v_1)\right|^{\alpha'}\right)+\left(2^j\right)^{\alpha-r}\left(1+\left|2^{-j}(x_2-v_2)\right|^{\alpha'}\right)\right]. \qquad \text{(A1.9)}$$

At $x=v$, *it follows that* $\forall k\le\lfloor\alpha\rfloor+1$,

$$\left|D_1^{k-r}D_2^r\Delta_j(x)\right|\le K\left[\left(2^j\right)^{\alpha-k+r}+\left(2^j\right)^{\alpha-r}\right] \qquad \text{(A1.10)}$$

Before computing $\left|f(x)-p_v(x)\right|$ *directly, we first need to check if the polynomial* p_v *(eqn.A1.6) has finite number of coefficients.*

<u>*Case 2.1*</u> $u\in U(v)$

For $u\in U(v)$, *we can choose* $v=\left(2^{-j}v_1,2^{-j}v_2\right)=u$, *so eqn.A1.7 becomes*

$$\left|\Delta_j(x)\right| \le K\left(2^j\right)^{\alpha} \int\int_{-\infty}^{+\infty} \frac{1}{1+\left(2^{-j}x_1-u_1\right)^q+\left(2^{-j}x_2-u_2\right)^q} \frac{du_1 du_2}{2^j}.$$

By the fast decay with $q=\alpha'+2$ *and at* $x=v$, *we can have*

$$\left|D_1^{k-r}D_2^r\ \Delta_j(x)\right| \le K\left[\left(2^j\right)^{\alpha-k+r}+\left(2^j\right)^{\alpha-r}\right],\quad \forall k \le \lfloor\alpha\rfloor+1,$$

from eqn.A1.10 directly.

Case 2.2 $u \notin U(v)$

Observing that $\left|2^{-j}x_1-u_1\right| \ge \delta/2$ *and* $\left|2^{-j}x_2-u_2\right| \ge \delta/2$ *for* $x \in U_{\delta/2}(v)$ *and* α *is not an integer so* $\alpha > \lfloor\alpha\rfloor$. *These guarantee a fast decay of* $\left|D_1^{k-r}D_2^r\ \psi_j^l(x)\right|$ *in eqn.A1.5 when* 2^j *goes to zero. At large scales* 2^j, *since* $\left|W_j^l f\right| \le \left\| f \right\| \left\| \psi^l \right\|$, *with the change of variables* $u'=2^{-j}x-u$ *in eqn.A1.5, we have*

$$\left|D_1^{k-r}D_2^r\ \Delta_j(v)\right| \le \frac{\left\| f \right\| \left\|\psi^l\right\|}{C_\psi} \left(\int\int_{-\infty}^{+\infty} \left|D_1^{k-r}D_2^r\ \psi^l(u')\right| du_1' du_2' \right) \int_{2^j}^{2^{j+1}} \frac{ds}{s^{\frac{3}{2}+k}}$$

and hence $\left|D_1^{k-r}D_2^r\ \Delta_j(x)\right| \le K\left[\left(2^j\right)^{\alpha-k+r}+\left(2^j\right)^{\alpha-r}\right],\quad \forall k \le \lfloor\alpha\rfloor+1.$

Together with eqn.A1.10 this proves that the polynomial p_v *defined in eqn.A1.6 has coefficients that are finite.*

Now, with the Littlewood-Paley decomposition, we compute directly

$$\left| f(x)-p_v(x)\right| = \left| \sum_{j\to-\infty}^{+\infty} \left(\Delta_j(x_1,x_2) - \sum_{k=0}^{\lfloor\alpha\rfloor} \frac{\left[\left((x_1-v_1)D_1+(x_2-v_2)D_2\right)^k \Delta_j\right](v_1,v_2)}{k!} \right) \right|$$

The sum over scales is divided into two at 2^J *such that* $2^J \ge \left|x-v\right| \ge 2^{J-1}$.

For $j \geq J$: we can use the classical Taylor theorem to bound the Taylor expansion of

$$\Delta_j, \qquad I = \sum_{j=J}^{+\infty} \left| \Delta_j(x_1, x_2) - \sum_{k=0}^{\lfloor \alpha \rfloor} \frac{\left[\left((x_1 - v_1) D_1 + (x_2 - v_2) D_2 \right)^k \Delta_j \right](v_1, v_2)}{k!} \right|$$

$$\leq \sum_{j=J}^{+\infty} \left| \frac{\sup_{u \in U(v)} \left[\left((x_1 - v_1) D_1 + (x_2 - v_2) D_2 \right)^{\lfloor \alpha \rfloor + 1} \Delta_j \right](u)}{(\lfloor \alpha \rfloor + 1)!} \right|$$

$$\leq \frac{1}{(\lfloor \alpha \rfloor + 1)!} \sum_{r=0}^{\lfloor \alpha \rfloor + 1} C_r^{\lfloor \alpha \rfloor + 1} (x_1 - v_1)^{\lfloor \alpha \rfloor + 1 - r} (x_2 - v_2)^r \sum_{j=J}^{+\infty} \left(\sup_{u \in U(v)} \left| D_1^{\lfloor \alpha \rfloor + 1 - r} D_2^r \Delta_j(u) \right| \right)$$

Inserting eqn.A1.9 yields

$$I \leq K \sum_{r=0}^{\lfloor \alpha \rfloor + 1} C_r^{\lfloor \alpha \rfloor + 1} (x_1 - v_1)^{\lfloor \alpha \rfloor + 1 - r} (x_2 - v_2)^r \sum_{j=J}^{+\infty} \left(\left(2^{-j}\right)^{\lfloor \alpha \rfloor + 1 - r - \alpha} \left| 2^{-j}(x_1 - v_1) \right|^{\alpha'} + \left(2^{-j}\right)^{r - \alpha} \left| 2^{-j}(x_2 - v_2) \right|^{\alpha'} \right)$$

and since $2^J \geq |x_i - v_i| \geq 2^{J-1}$ for $i = 1, 2$, we get $I \leq K \left(|x_1 - v_1|^2 + |x_2 - v_2|^2 \right)^{\frac{\alpha}{2}}$.

For $j < J$:

$$II = \sum_{j \to -\infty}^{J-1} \left| \left(\Delta_j(x_1, x_2) - \sum_{k=0}^{\lfloor \alpha \rfloor} \frac{\left[\left((x_1 - v_1) D_1 + (x_2 - v_2) D_2 \right)^k \Delta_j \right](v_1, v_2)}{k!} \right) \right|$$

$$\leq \sum_{j \to -\infty}^{J-1} \left| \Delta_j(x_1, x_2) \right| + \sum_{j \to -\infty}^{J-1} \left| \sum_{k=0}^{\lfloor \alpha \rfloor} \frac{\left[\left((x_1 - v_1) D_1 + (x_2 - v_2) D_2 \right)^k \Delta_j \right](v_1, v_2)}{k!} \right|$$

By eqn.A1.8, the first term

$$\sum_{j \to -\infty}^{J-1} \left| \Delta_j(x_1, x_2) \right| \leq K \sum_{j \to -\infty}^{J-1} \left(2^j\right)^{\alpha} \left[2 + 2^{\alpha'} \left(\left| 2^{-j}(x_1 - v_1) \right|^{\alpha'} + \left| 2^{-j}(x_2 - v_2) \right|^{\alpha'} \right) \right]$$

$$\leq K \left(2^{J\alpha} + 2^{J(\alpha - \alpha')} \left(|x_1 - v_1|^{\alpha'} + |x_2 - v_2|^{\alpha'} \right) \right)$$

Also, by eqn.A1.10, the second term is bounded by

$$K \sum_{j \to -\infty}^{J-1} \sum_{r=0}^{\lfloor \alpha \rfloor} \frac{C_r^{\lfloor \alpha \rfloor} (x_1 - v_1)^{\lfloor \alpha \rfloor - r} (x_2 - v_2)^r}{r!} \left(\left(2^j\right)^{\alpha - \lfloor \alpha \rfloor + r} + \left(2^j\right)^{\alpha - r} \right)$$

$$\leq K \sum_{r=0}^{\lfloor \alpha \rfloor} \frac{C_r^{\lfloor \alpha \rfloor} (x_1 - v_1)^{\lfloor \alpha \rfloor - r} (x_2 - v_2)^r}{r!} \left(2^{J(\alpha - \lfloor \alpha \rfloor + r)} + 2^{J(\alpha - r)} \right)$$

and since $2^J \geq |x_i - v_i| \geq 2^{J-1}$ *for* $i = 1, 2$, *we get* $II \leq K\left(|x_1 - v_1|^2 + |x_2 - v_2|^2\right)^{\frac{\alpha}{2}}$.

As a result, $|f(x_1, x_2) - f(v_1, v_2)| \leq I + II \leq K\left(|x_1 - v_1|^2 + |x_2 - v_2|^2\right)^{\frac{\alpha}{2}}$, *which proves that* f *is uniformly Lipschitz* α *in* $U_\delta(v)$.

Appendix II

Proof of the necessary condition of theorem 4.1

For $l=1,2$, $\left|d_{l,n}^{(j)}\right| = \left|\int_{-\infty}^{+\infty} f(t)2^{j/2}\psi_l\left(2^j t-n\right)dt\right|$

$$= \left|\int_{-\infty}^{+\infty}\left(f(t)-p_v(t)\right)2^{j/2}\psi_l\left(2^j t-n\right)dt\right|, \qquad \text{by eqn.4.13,}$$

$$\le \int_{-\infty}^{+\infty} K\left|t-v\right|^{\alpha} 2^{j/2}\left|\psi_l\left(2^j t-n\right)\right|dt, \qquad \text{by definition 4.2.}$$

With the change of variable, $x=2^j t-n$, *we have*

$$\left|d_{l,n}^{(j)}\right| \le \int_{-\infty}^{+\infty} K\left|2^{-j}x+2^{-j}n-v\right|^{\alpha} 2^{j/2}2^{-j}\left|\psi_l(x)\right|dx$$

$$= 2^{-j/2}\int_{-\infty}^{+\infty} K\left|2^{-j}x+2^{-j}n-v\right|^{\alpha}\left|\psi_l(x)\right|dx.$$

By Cauchy Schwarz Inequality, $|a+b|^r \le 2^r\left(|a|^r+|b|^r\right)$, *we have*

$$\left|d_{l,n}^{(j)}\right| \le K\cdot 2^{\alpha}\cdot 2^{-j/2}\left(2^{-j\alpha}\int_{-\infty}^{+\infty}|x|^{\alpha}\left|\psi_l(x)\right|dx+\left|2^{-j}n-v\right|^{\alpha}\int_{-\infty}^{+\infty}\left|\psi_l(x)\right|dx\right)$$

$$= K2^{\alpha}\left(2^{-j}\right)^{\alpha+1/2}\left(\int_{-\infty}^{+\infty}|x|^{\alpha}\left|\psi_l(x)\right|dx+\left|n-2^j v\right|^{\alpha}\int_{-\infty}^{+\infty}\left|\psi_l(x)\right|dx\right)$$

$$= K2^{\alpha}\left(2^{-j}\right)^{\alpha+1/2}\int_{-\infty}^{+\infty}|x|^{\alpha}\left|\psi_l(x)\right|dx\left(1+\frac{\int_{-\infty}^{+\infty}\left|\psi_l(x)\right|dx}{\int_{-\infty}^{+\infty}|x|^{\alpha}\left|\psi_l(x)\right|dx}\left|n-2^j v\right|^{\alpha}\right)$$

$$\le A\left(2^{-j}\right)^{\alpha+1/2}\left(1+\left|n-2^j v\right|^{\alpha}\right). \qquad \text{(A2.1)}$$

For $n\in[a,b]$, *we can choose* $n=2^j v$, *we have* $\left|d_{l,n}^{(j)}\right| \le A\left(2^{-j}\right)^{\alpha+1/2}$.

Proof of the sufficient condition of theorem 4.1

As the discrete wavelet decomposition formula is given as the following

$$f(t) = \sum_n\left(\sum_{l=1}^{2} c_{l,n}^{(J)}2^{J/2}\phi_l\left(2^J t-n\right)+\sum_{j=J}^{-1}\sum_{l=1}^{2} d_{l,n}^{(j)}2^{j/2}\psi_l\left(2^j t-n\right)\right)$$

Let $\Delta_j(f) = \sum_n \sum_{l=1}^{2} d_{l,n}^{(j)} 2^{j/2} \psi_l\left(2^j t - n\right)$. *If* f *is* n *times differentiable at* v, *then* p_v *corresponds to the Taylor polynomial but this is not necessarily true. We shall first prove that* $\sum_{j=-\infty}^{+\infty} \Delta_j^{(k)}(f)(v)$ *is finite by getting upper bounds on* $\left|\Delta_j^{(k)}(f)(t)\right|$. *These sums may be thought of as a generalization of pointwise derivatives.*

To simplify the notation, we denote by K *a generic constant which may change value from one line to the next but does not depend on* j *and* t. *The hypothesis eqn.4.14 and the asymptotic decay condition eqn.4.10 imply that*

$$\begin{aligned} \left|\Delta_j(f)\right| &= \left|\sum_n \sum_{l=1}^{2} d_{l,n}^{(j)} 2^{j/2} \psi_l\left(2^j t - n\right)\right| \\ &\le \sum_n \sum_{l=1}^{2} \left|d_{l,n}^{(j)}\right| 2^{j/2} \left|\psi_l\left(2^j t - n\right)\right| \\ &\le A\left(2^{-j}\right)^{\alpha+1/2} \sum_n \sum_{l=1}^{2} \left|\psi_l\left(2^j t - n\right)\right| \\ &= C\left(2^{-j}\right)^{\alpha+1/2}. \end{aligned} \tag{A2.2}$$

Using the localization of wavelets and of their derivatives, we also get

$$\left|\Delta_j^{(k)}(f)\right| \le C\left(2^j\right)^k \left(2^{-j}\right)^{\alpha+1/2}$$

and so

$$\left|\Delta_j^{(i)}(f)\right| \le C 2^j \left(2^{-j}\right)^{\alpha+1/2}. \tag{A2.3}$$

where C *is a generic constant which changes from line to line.*

Let J *be such that* $2^{-J} \le |t - v| \le 2 \cdot 2^{-J}$.

Then $$|f(t) - f(v)| \le \sum_{j \le J} \left|\Delta_j(f)(t) - \Delta_j(f)(v)\right| + \sum_{j > J} \left(\left|\Delta_j(f)(t)\right| + \left|\Delta_j(f)(v)\right|\right).$$

Using the mean value theorem and the results of eqn.A2.2 and A2.3, the first sum is bounded by $C \sum_{j \le J} |t - v| 2^j \left(2^{-j}\right)^{\alpha+1/2} \le C |t - v|^{\alpha}$.

and the second sum is bounded by $C\sum_{j>J}\left(2^{-J}\right)^{\alpha+1/2} \le C\left|t-v\right|^{\alpha}$.

Therefore, f *is uniformly Lipschitz* α *over* $[a,b]$.

Appendix III

Proof of the necessary condition of theorem 4.2

By eqn.A2.1, eqn.4.15 is proved to be true for f *is Lipschitz* α *at* v.

Proof of the sufficient condition of theorem 4.2

$$
\begin{aligned}
|\Delta_j(f)| &= \left|\sum_n \sum_{l=1}^{2} d_{l,n}^{(j)} 2^{j/2} \psi_l\left(2^j t - n\right)\right| \\
&\le \sum_n \sum_{l=1}^{2} \left|d_{l,n}^{(j)}\right| 2^{j/2} \left|\psi_l\left(2^j t - n\right)\right| \\
&\le \sum_n \sum_{l=1}^{2} A\left(2^{-j}\right)^{\alpha+1/2} \left(1+\left|n-2^j v\right|^{\alpha'}\right) 2^{j/2} \left|\psi_l\left(2^j t - n\right)\right|, \quad \textit{by eqn.4.16,} \\
&= A\left(2^{-j}\right)^{\alpha} \sum_n \sum_{l=1}^{2} \left(1+\left|n-2^j v\right|^{\alpha'}\right) \left|\psi_l\left(2^j t - n\right)\right| \\
&\le A\left(2^{-j}\right)^{\alpha} \sum_n \sum_{l=1}^{2} \left(1+\left|n-2^j v\right|^{\alpha'}\right) \frac{C_{l,q}}{1+\left|2^j t - n\right|^{q}}, \quad \textit{by eqn.4.10,} \\
&= C\left(2^{-j}\right)^{\alpha} \sum_n \sum_{l=1}^{2} \frac{1+\left|n-2^j v\right|^{\alpha'}}{1+\left|2^j t - n\right|^{q}}.
\end{aligned}
$$

Since $|u-v|^{\alpha'} \le 2^{\alpha'}\left(|u-t|^{\alpha'} + |t-v|^{\alpha'}\right)$, *replacing* u *with* $2^{-j}n$ *we have*

$$
|\Delta_j(f)| \le C\left(2^{-j}\right)^{\alpha} \sum_n \sum_{l=1}^{2} \frac{1+\left|\frac{2^{-j}n-t}{2^{-j}}\right|^{\alpha'} + \left|\frac{t-v}{2^{-j}}\right|^{\alpha'}}{1+\left|\frac{t-2^{-j}n}{2^{-j}}\right|^{q}}.
$$

With the change of variable $n' = \dfrac{2^{-j}n - t}{2^{-j}}$ *yields*

$$
|\Delta_j(f)| \le C\left(2^{-j}\right)^{\alpha} \sum_{n'} \sum_{l=1}^{2} \frac{1+|n'|^{\alpha'} + \left|\frac{t-v}{2^{-j}}\right|^{\alpha'}}{1+|n'|^{q}}.
$$

Choosing $q = \alpha' + 2$ *yields*

$$|\Delta_j(f)| \leq C(2^{-j})^{\alpha}\left(1 + + \left|\frac{t-v}{2^{-j}}\right|^{\alpha'}\right). \qquad \text{(A3.1)}$$

The same derivation applied to the derivatives of $\Delta_j(f)$

$$\forall k \leq \lfloor \alpha \rfloor + 1, \qquad |\Delta_j^{(k)}(f)| \leq K(2^{-j})^{\alpha-k}\left(1 + \left|\frac{v-t}{2^j}\right|^{\alpha'}\right).$$

At $t = v$ *it follows that*

$$\forall k \leq \lfloor \alpha \rfloor, \qquad |\Delta_j^{(k)}(f)| \leq K(2^{-j})^{\alpha-k}. \qquad \text{(A3.2)}$$

This guarantees a fast decay of $|\Delta_j^{(k)}(f)|$ *when* 2^{-j} *goes to zero, because* α *is not an integer so* $\alpha > \lfloor \alpha \rfloor$. *At large scales* 2^{-j}, *since* $|d_{l,n}^{(j)}| \leq \|f\|\|\psi_l\|$,

$$\begin{aligned} |\Delta_j(f)| &= \left|\sum_n \sum_{l=1}^{2} d_{l,n}^{(j)} 2^{j/2} \psi_l(2^j t - n)\right| \\ &\leq \sum_n \sum_{l=1}^{2} |d_{l,n}^{(j)}| 2^{j/2} |\psi_l(2^j t - n)| \\ &= \|f\|\|\psi_1\| \sum_n 2^{j/2} |\psi_1(2^j t - n)| + \|f\|\|\psi_2\| \sum_n 2^{j/2} |\psi_2(2^j t - n)|. \end{aligned}$$

Together with eqn.A3.2 this proves that the polynomial $p_v(t) = \sum_{k=0}^{\lfloor \alpha \rfloor} \sum_{j=-\infty}^{+\infty} \Delta_j^{(k)}(v) \frac{(t-v)^k}{k!}$

has coefficients that are finite.

Now, with the Littlewood-Paley decomposition we compute

$$|f(t) - p_v(t)| = \left|\sum_{j=-\infty}^{+\infty} \left(\Delta_j(t) - \sum_{k=0}^{\lfloor \alpha \rfloor} \Delta_j^{(k)}(v) \frac{(t-v)^k}{k!}\right)\right|.$$

Again, the sum over scales is divided in two at 2^{-j} *such that* $2^{-j} \leq |t-v| \leq 2 \cdot 2^{-j}$.

For $j \geq J$, we can use the classical Taylor theorem to bound the Taylor expansion of

Δ_j,

$$I = \sum_{j=J}^{+\infty} \left| \Delta_j(t) - \sum_{k=0}^{\lfloor \alpha \rfloor} \Delta_j^{(k)}(v) \frac{(t-v)^k}{k!} \right|$$

$$\leq \sum_{j=J}^{+\infty} \frac{(t-v)^{\lfloor \alpha \rfloor + 1}}{(\lfloor \alpha \rfloor + 1)!} \sup_{h \in [t,v]} \left| \Delta_j^{\lfloor \alpha \rfloor + 1}(h) \right|.$$

Inserting eqn.A3.1 yields

$$I \leq K |t-v|^{\lfloor \alpha \rfloor + 1} \sum_{j=J}^{+\infty} \left(2^{-j}\right)^{\alpha - \lfloor \alpha \rfloor - 1} \left| 2^j (t-v) \right|^{\alpha'}$$

and since $2^{-J} \leq |t-v| \leq 2 \cdot 2^{-J}$, we get $I \leq C |t-v|^{\alpha}$.

For $j < J$,

$$II = \sum_{j=-\infty}^{J-1} \left| \Delta_j(t) - \sum_{k=0}^{\lfloor \alpha \rfloor} \Delta_j^{(k)}(v) \frac{(t-v)^k}{k!} \right|$$

$$\leq K \sum_{j=-\infty}^{J-1} \left(\left(2^{-j}\right)^{\alpha} \left(1 + \left| 2^j (t-v) \right|^{\alpha'} \right) + \sum_{k=0}^{\lfloor \alpha \rfloor} \frac{(t-v)^k}{k!} \left(2^{-j}\right)^{\alpha - k} \right)$$

$$\leq K \left(\left(2^{-J}\right)^{\alpha} + \left(2^{-J}\right)^{\alpha - \alpha'} |t-v|^{\alpha'} + \sum_{k=0}^{\lfloor \alpha \rfloor} \frac{(t-v)^k}{k!} \left(2^{-J}\right)^{\alpha - k} \right)$$

and since $2^{-J} \leq |t-v| \leq 2 \cdot 2^{-J}$, we get $II \leq K |v-t|^{\alpha}$. As a result

$$|f(t) - p_v(t)| \leq I + II \leq K |v-t|^{\alpha}$$

which proves that f is Lipschitz α at v.

Appendix IV

Wavelet Filter Definitions

1. Daubechies's biorthogonal 9/3*:

$$h[n] = (1/2 \quad 1/4 \quad 12)$$

$$g[n] = (3/128 \quad -3/64 \quad -1/8 \quad 19/64 \quad 45/64 \quad 19/64 \quad -1/8 \quad -3/64 \quad 3/128)$$

Multiwavelet Filter Definitions

Prefilters and Postfilters

1. Identity:

$$Q_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

2. Interpolation:

$$Q_0 = \begin{pmatrix} 3/8\sqrt{6} & 5/4\sqrt{6} \\ 0 & 0 \end{pmatrix}, \qquad Q_1 = \begin{pmatrix} 3/8\sqrt{6} & 0 \\ 1/\sqrt{3} & 0 \end{pmatrix}.$$

$$P_0 = \begin{pmatrix} 0 & 0 \\ 4\sqrt{6}/5 & -3\sqrt{3}/10 \end{pmatrix}, \qquad P_0 = \begin{pmatrix} 0 & \sqrt{3} \\ 0 & -3\sqrt{3}/10 \end{pmatrix}.$$

3. Minimax [106]:

$$Q_0 = \begin{pmatrix} 2\sqrt{2} & -\sqrt{2} \\ 1 & 0 \end{pmatrix}.$$

$$P_0 = \begin{pmatrix} 0 & 1 \\ -1/\sqrt{2} & 2 \end{pmatrix}.$$

4. Minimal Repeated Signal [49]:

defined by the vector $\gamma_o = \begin{pmatrix} \sqrt{2} & 1 \end{pmatrix}^T$, which is the first eigenvalue of $\tilde{H}(0)$.

$$S_{0,k} = f_{k+1} \begin{pmatrix} \sqrt{2} \\ 1 \end{pmatrix}.$$

5. Plonka* [43]:

$$Q_0 = \begin{pmatrix} \sqrt{6}/6 & \sqrt{6}/6 \\ \sqrt{3}/3 & 0 \end{pmatrix}.$$

$$P_0 = \begin{pmatrix} 0 & 1 \\ -1/\sqrt{2} & 2 \end{pmatrix}.$$

6. Xia *et. al* [57]:

$$Q_0 = \begin{pmatrix} 2+\sqrt{2}/10 & 2-\sqrt{2}/10 \\ \sqrt{2}+3/20 & \sqrt{2}-3/20 \end{pmatrix}.$$

7. Xia* [59]:

$$Q_0 = \begin{pmatrix} \sqrt{6}-1/2 & \sqrt{6}+1/2 \\ \sqrt{3}/2 & \sqrt{3}/2 \end{pmatrix}.$$

8. Yang *et. al* [60]:

$$Q_0 = \begin{pmatrix} \sqrt{6}/9 & 2\sqrt{6}/9 \\ -\sqrt{3}/9 & 4\sqrt{3}/9 \end{pmatrix}.$$

Multifilters

1. GHM* [21]:

$$H_0 = \begin{pmatrix} 3/10 & 2\sqrt{2}/5 \\ -\sqrt{2}/40 & -3/20 \end{pmatrix}, \qquad H_1 = \begin{pmatrix} 3/10 & 0 \\ 9\sqrt{2}/40 & 1/2 \end{pmatrix},$$

$$H_2 = \begin{pmatrix} 0 & 0 \\ 9\sqrt{2}/40 & -3/20 \end{pmatrix}, \qquad H_3 = \begin{pmatrix} 0 & 0 \\ -\sqrt{2}/40 & 0 \end{pmatrix}.$$

$$G_0 = \begin{pmatrix} -\sqrt{2}/40 & -3/20 \\ -1/20 & -3\sqrt{2}/20 \end{pmatrix}, \qquad G_1 = \begin{pmatrix} 9\sqrt{2}/40 & -1/2 \\ 9/20 & 0 \end{pmatrix},$$

$$G_2 = \begin{pmatrix} 9\sqrt{2}/40 & -3/20 \\ -9/20 & 3\sqrt{2}/20 \end{pmatrix}, \qquad G_3 = \begin{pmatrix} -\sqrt{2}/40 & 0 \\ 1/20 & 0 \end{pmatrix}.$$

2. Chui-Lian* [8]:

$$H_0 = \frac{1}{2}\begin{pmatrix} 1/2 & 1/2 \\ -\sqrt{7}/4 & -\sqrt{7}/4 \end{pmatrix}, \qquad H_1 = \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & 1/2 \end{pmatrix},$$

$$H_2 = \frac{1}{2}\begin{pmatrix} 1/2 & -1/2 \\ \sqrt{7}/4 & -\sqrt{7}/4 \end{pmatrix}.$$

$$G_0 = \frac{1}{2}\begin{pmatrix} -1/2 & -1/2 \\ 1/4 & 1/4 \end{pmatrix}, \qquad G_1 = \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & \sqrt{7}/2 \end{pmatrix},$$

$$G_2 = \frac{1}{2}\begin{pmatrix} -1/2 & 1/2 \\ -1/4 & 1/4 \end{pmatrix}.$$

* : simulation results are shown in this thesis

Bibliography

Academic Journal Papers:

[1] Berman, Z. and Baras J. S. "Properties of the multiscale maxima and zero-crossings representations". *IEEE Trans. on Signal Processing*, 41(12):3216-3231 (1993)

[2] Bui, T. D. and Chen, G. "Translation-Invariant Denoising Using Multiwavelets". *IEEE Trans. on Signal Processing*, 46(12):3414-3420 (1998)

[3] Calderón, A. P. "Intermediate spaces and interpolation, the complex method". *Stud. Math.*, 24:113-190 (1964)

[4] Canny, J. "A Computational Approach to Edge Detection". *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 8(5):961-1005 (1986)

[5] Carey, W. K., Chuang, D. B. and Hemami, S. S. "Regularity-Preserving Image Interpolation". *IEEE Trans. on Image Processing*, 8(9):1293-1297 (1999)

[6] Chan, T. C.-L. and Lun, D. P.-K. "Adaptive Thresholding for Noisy MPEG-4 Still Texture Image Coding". *to appear in IEEE Trans. on Image Processing*

[7] Chang, G., Yu, B. and Vetterli, M. "Adaptive wavelet thresholding for image denoising and compression". *IEEE Trans. on Image Processing*, 9(9):1532-1546 (2000)

[8] Chui, C. K. and Lian, J. "A Study of Orthonormal Multi-wavelets". *Applied Numerical Mathematics*, 20:273-298 (1996)

[9] Cohen, I., Daubechies, I. And Vial, P. "Wavelet bases on the interval and fast algorithms". *J. of Applied and Computational Harmonic Analysis*, 1:54-81 (1993)

[10] Cvetkovic, Z. and Vetterli, M. "Discrete-time wavelet extrema representation: design and consistent reconstruction". *IEEE Trans. on Signal Processing*, 43(3):681-693 (1995)

[11] Daubechies, I. "Orthonormal bases of compactly supported wavelets". *Commun. on Pure and Appl. Math.*, 41(6):909-996 (1988)

[12] Daubechies, I. "The wavelet transform, time-frequency localization and signal analysis". *IEEE Trans. on Information Theory*, 36(5):961-1005 (1990)

[13] Donoho, D. L. and Johnstone, I. M. "Ideal spatial adaptiation via wavelet shrinkage". *Biometrika*, 81:425-455 (1994)

[14] Donoho, D. L. "De-noising by soft-thresholding". *IEEE Trans. on Information Theory*, 41(5):613-627 (1995)

[15] Donoho, D. L., Johnstone, I. M., Kerkyacharian, G. and Picard, D. "Wavelet shrinkage: Asymptopia?". *J. R. Stat. B.*, 57:301-369 (1995)

[16] Donoho, D. L., Johnstone, I. M. "Adapting to unknown smoothness via wavelet shrinkage". *J. of the American Statistical Association*, 90(12):1200-1224 (1995)

[17] Donovan, G. C., Geronimo, J. S., Hardin, D. P. and Massopust, P. R.. "Construction of orthogonal wavelets using fractal interpolation". SIAM Math. Anal., 27:1158–1192 (1996)

[18] Downie, T. R. and Silverman, B. W. "The Discrete Multiple Wavelet Transform and Thresholding Methods". *IEEE Trans. on Signal Processing*, 46(9):2558-2561 (1998)

[19] Duffin, R. J. and Schaeffer, A. C. "A class of nonharmonic Fourier series". *Trans. Amer. Math. Soc.*, 72:341-366 (1952)

[20] Gabor, D. "Theory of communication". *J. IEE*. 93:429-457 (1946)

[21] Geronimo, J. S., Hardin, D. P. and Massopust, P. R. "Fractal Functions and Wavelet Expansions Based on Several Scaling Functions". *J. of Approximation Theory*, 78:373-401 (1994)

[22] Goodman, T. N. T., Lee, S. L. and Tang, W. S. "Wavelets in Wandering Subspaces". *Trans. of the American Mathematical Society*, 338(2):639-654 (1993)

[23] Goodman, T. N. T. and Lee, S. L. "Wavelets of Multiplicity r". *Trans. of the American Mathematical Society*, 342(1):307-324 (1994)

[24] Goodman, T. N. T. "Interpolatory Hermite Spline Wavelets". *J. of Approximation Theory*, 78:174-189 (1994)

[25] Gopinath, R. A., Odegard, J. E. and Burrus, C. S. "Optimal Wavelet Representation of Signals and the Wavelet Sampling Theorem". *IEEE Trans. on Circuits and Systems*, 41(4):262-277 (1994)

[26] Grossmann, A. and Morlet, J. "Decomposition of Hardy Functions into square integrable wavelets of constant shape". *SIAM J. Math. Anal.*, 15(4):723-736, (1984)

[27] Gröchenig, K. "Irregular sampling of wavelet and short-time Fourier transforms". *Constr. Approx.*, 9:287-297 (1993)

[28] Haar, A. "Zur theorie der orthogonalen funktionensysteme". *Math. Annal.*, 69:331-371 (1910)

[29] Hardin, D. P. and Roach, D. W. "Multiwavelet prefilters I: Orthogonal prefilters preserving approximation order $p \leq 2$". *IEEE Trans. on Circuit and Systems II: Analog and Digital Signal Processing*, 45(8):1106-1112 (1998)

[30] Ho, Y.-F., Hsung, T.-C., Lun, D. P.-K. and Tam, P. K.-S. "Regularity Scalable Wavelet Image Coding Based on Singularity Detection". *IEEE Trans. on Image Processing*. (to be submitted)

[31] Ho, Y.-F., Hsung, T.-C., Lun, D. P.-K. and Tam P. K.-S. "Multiwavelet Denoising Using Singularity Detection," *IEEE Trans. on Image Processing*. (to be submitted).

[32] Hsung, T.-C., Lun, D. P.-K. and Siu, W.-C. "Denoising by Singularity Detection". *IEEE Trans. on Signal Processing*, 47(11):3139-3144 (1999)

[33] Jansen, M., Malfait, M. and Bultheel, A. "Generalized Cross Validation for wavelet thresholding". *Signal Processing*, 56(1):33-44 (1997)

[34] Jiang, Q. T. "Orthogonal Multiwavelets with Optimum Time-Frequency Resolution". *IEEE Trans. on Signal Processing*, 46(4):830-844 (1998)

[35] Lebrun J. and Vetterli, M. "Balanced Multiwavelets Theory and Design". *IEEE Trans. on Signal Processing*, 46(4):1119-1125 (1998)

[36] Lewis, A. S. and Knowles, G. "Image compression using the 2-D wavelet transform". *IEEE Trans. on Image Processing*, 1(4):244-250, (1992)

[37] Mallat, S. "Multiresolution approximations and wavelet orthonormal bases of $L^2(\Re)$". *Trans. Amer. Math. Soc.*, 315(9):69-87 (1989)

[38] Mallat, S. "A theory for multiresolution signal decomposition: the wavelet representation". *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 11(7):674-693 (1989)

[39] Mallat, S. and Zhong, S. "Characterization of signals from multiscale edges". *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 14(7):710-732 (1992)

[40] Mallat, S. and Hwang, W. "Singularity Detection and Processing with Wavelets". *IEEE Trans. on Information Theory*, 38(3):617-643 (1992)

[41] Mintzer, F. "Filters for distortion-free two-band multirate filter banks". *IEEE Trans. on Acoustics, Speech and Signal Processing*, 33(3):626-630 (1985)

[42] Nason, G. P. "Wavelet shrinkage using cross-validation". *J. R. Stat. Soc. Ser. B*, 58:463-479 (1996)

[43] Plonka, G. "Approximation properties of multiscaling functions: A Fourier approach". *Rostocker Mathematische Kolloquium*, 49: 115-126 (1995)

[44] Rioul, O. and Duhamel, P. "Fast algorithms for discrete and continuous wavelet transforms". *IEEE Trans. on Information Theory*, 38(3):569-586 (1992)

[45] Said, A. and Pearlman, W. "A New, Fast, and Efficient Image Codec Based on Set Partitioning in Hierarchical Trees". *IEEE Trans. on Circuits and Systems for Video Technology*, 6(3):243-250 (1996)

[46] Simencelli, E. P., Freeman, W. T., Adelson, E. H. and Heeger, D. J. "Shiftable Multiscale Transforms". *IEEE Trans. on Information Theory*, 38(2):587-607 (1992)

[47] Shapiro, J. M. "Embedded Image Coding Using Zerotrees of Wavelet Coefficients". *IEEE Trans. on Signal Processing*, 41(12)3445-3462 (1993)

[48] Strang, G. and Strela, V. "Short Wavelets and Matrix Dilation Equations". *IEEE Trans. on Signal Processing*, 43(1):108-115 (1995)

[49] Strela, V., Heller, P. N., Strang, G., Topiwala, P. and Heil, C. "The Application of Multiwavelet Filterbanks to Image Processing". *IEEE Trans. on Signal Processing*, 8(4):548-563 (1999)

[50] Taubman, D. "High Performance Scalable Image Compression with EBCOT". *IEEE Trans. on Image Processing*, 9(7):1158-1170 (2000)

[51] Vaidyanathan, P. V. "Quadrature mirror filter banks, M-band extensions and perfect reconstruction techniques". *IEEE ASSP Magazine*, 4(3):4-20 (1987)

[52] Vetterli, M. "Filter banks allowing perfect reconstruction". *Signal Processing*, 10(3):219-244 (1986)

[53] Vetterli, M. and Herley, C. "Wavelets and Filter Banks: theory and design". *IEEE Trans. on Signal Processing*, 40(9):2207-2232 (1992)

[54] Ville, J. "Theorie et applications de la notion de signal analytique". *Cables et Transm.*, 2A(1): 61-74 (1948)

[55] Wigner, E. P. "On the quantum correction for thermodynamic equilibrium". *Phys. Rev.* 40:749-759 (1932)

[56] Xia, X.-G. and Suter, B. W. "Vector-Valued Wavelets and Vector Filter Banks". *IEEE Trans. on Signal Processing*, 44(3):508-518 (1996)

[57] Xia, X.-G., Geronimo, J. S., Hardin, D. P. and Suter, B. W. "Design of Prefilters for Discrete Multiwavelet Transform". *IEEE Trans. on Signal Processing*, 44(1):25-35 (1996)

[58] Xia, X.-G. and Suter, B. W. "Vector-Valued Wavelets and Vector Filter Banks". *IEEE Trans. on Signal Processing*, 44(3):508-518 (1996)

[59] Xia, X.-G. "A New Prefilter Design for Discrete Multiwavelet Transforms". *IEEE Trans. on Signal Processing*, 46(6):1558-1570 (1998)

[60] Yang, X. X., Jiao, L. C. and Zhang, J. K. "Design of orthogonal prefilter with the Strang-Fix condition". *Electronics Letters*, 35 (2):117-119, 21st Jan. (1999)

Books and Book Chapters:

[61] Antoniadis, A. and Oppenheim, G. *Wavelets and Statistics*. Springer (1995)

[62] Akansu, A. and Haddad, R. *Multiresolution Signal Decomposition*. Academic Press (1993)

[63] Akansu, A. and Smith, M. J. *Subband and Wavelet Transform*. Kluwer (1995)

[64] Carmona, R. "Extrema reconstruction and spline smoothing: variations on an algorithm of Mallat and Zhong". In Antoniadis, A. and Oppenheim, G., eds., *Wavelets and Statistics*. Springer-Verlag, Berlin, pp.96-108 (1995)

[65] Coifman, R. R. and Donoho, D. L. "Translation invariant de-noising". In *Wavelets and Statistics*, Springer Lecture Notes in Statistics 103. pp.125-150 (1994)

[66] Chui, C. K. *An Introduction to Wavelets*. Academic, New York (1992)

[67] Daubechies, I. *Ten Lectures on Wavelets*. SIAM Philadelphia: PA (1992)

[68] Gopinath, R. A. and Burrus, C. S. "Wavelet Transforms and Filter Banks". In Chui, C. K., eds., *Wavelet: A Tutorial in Theory and Applications*. Academic Press, San Diego (1992)

[69] Held, G. and Marshall, T. R. *Data and Image Compression: Tools and Techniques*. John Wiley (1996)

[70] Heller, P. N. and Wells, R. O., Jr. "Spectral Theory of Multiresolution Operators and Applications". In Chui, C. K., eds., *Wavelets: Theory, Algorithms, and Applications*. Academic Press, San Diego (1994)

[71] Holschneider, M., Kronland-Martinet, R., Morlet, J. and Tchamitchian, P. "Wavelets, Time-Frequency Methods and Phase Space". In chapter, *Real-Time Algorithm for Signal Analysis with the Help of the Wavelet Transform*. Springer-Verlag, Berlin, pp.289-297 (1989)

[72] Holschneider, M. and Tchamitchian, P. "Regularite Locale de la Fonction Non-Differentiable de Riemann". In Lemarie, P. G., eds., *Les ondelettes en 1989: Seminarie d'analyse harmonique*. Universite de Paris-Sud, Orsay. Springer-Verlag, Berlin, New York (1990)

[73] Hsung, T. C. *Wavelet Transform and Its Applications to Image Processing*. PhD Dissertation, The Hong Kong Polytechnic University, (1998)

[74] Jaffard, S. "Regularity analysis of functions and random processes using wavelets". In Byrnes, J. S. et al., eds., *Wavelet and Their Applications*, Kluwer Academic Publishers, Netherlands, pp.51-82 (1994)

[75] Jaffard, S. *Wavelet Methods for Pointwise Regularity and Local Oscillations of Functions*. American Mathematical Society (1996)

[76] Mallat, S. *A Wavelet Tour of Signal Processing*. 2nd Ed., Academic Press, New York (1999)

[77] Meyer, Y. *Wavelets: Algorithms and Applications*. In Ryan, R. D., trans. rev., SIAM (1995)

[78] Rabbani, M. and Jones, P. W. *Digital Image Compression Techniques*. SPIE Optical Engineeering Press, Volume TT7 (1991)

[79] Sayood, K. *Introduction to Data Compression*. Morgan Kaufmann Publishers, Inc. (1996)

[80] Shapiro, H. S. "The generalized modulus of continuity and wavelets". In Byrnes, J. S. et al., eds., *Wavelets and Their Applications*. Kluwer Academic Publishers, Netherlands, pp.83-93 (1994)

[81] Strang, G. and Nguyen, T. *Wavelet and Filter Banks*. Wellesley-Cambridge Press, Boston (1996)

[82] Topiwala, P. N. *Wavelet Image And Video Compression*. Kluwer Academic Publishers (1998)

[83] Vaidyanathan, P. P. *Multirate Systems and Filter Banks*. Prentice Hall PTR, Engle-wood Cliffs (1993)

[84] Vetterli, M. and Kovacevic, J. *Wavelets and Subband Coding*. Prentice-Hall, Englewood Cliffs, NJ (1995)

Conference Papers:

[85] Croisier, A., Esteban, D. and Galand, C. "Perfect channel splitting by use of interpolation/-decimation/tree decomposition techniques". *Int. Conf. on Info. Sciences and Systems*, 1:443-446, Patras, Greece (1976)

[86] Ho, Y.-F., Hsung, T.-C. and Lun, D. P.-K. "Adaptive Regularity Scalable Wavelet Image Coding". *Proc. of IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, 1:3493-3496 (2002)

[87] Ho, Y.-F., Hsung, T.-C., Lun, D. P.-K. and Tam, P. K.-S. "Multiwavelet Denoising Using Singularity Detection," *IEEE Int. Conf. on Acoustics, Speech and Signal Processing 2003*. (submitted).

[88] Hsung, T.-C. and Lun, D. P.-K. "Denoising by singularity rejection". *Proc. of Int. Symp. on Circuits and Systems*, 1:205-208 (1997)

[89] Hsung, T.-C., Chan, T. C.-L., Lun, D. P.-K. and Feng, D. D. "Embedded Singularity Detection Zerotree Wavelet Coding". *Proc. of Int. Conf. on Image Processing*, 2:274-278 (1999)

[90] Kim, W. and Li, C.-C. "A study on Preconditioning Multiwavelet Systems for Image Compression". *Proc. of 2nd Int. Conf. on Wavelet Analysis and Its Applications*, 1:22-36 (2001)

[91] Law, N.-F. and Siu, W.-C. "Progressive Image Coding based on Visually Important Features". *Proc. of Int. Conf. on Image Processing*, 2:362-366 (1999)

[92] Lebrun J. and Vetterli, M. "Balanced Multiwavelets". *IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, 3:2463-2476 (1997)

[93] Ling, W. "Orthogonal Multiwavelets Transform for Image Denoising". *Proc. 5th Int. Conf. on Signal Processing*, 2:987-991 (2000)

[94] Liu, N. "Scalable Wavelet Image Coding With Photoclinometry Criterion". *Proc. of Int. Conf. on Information, Communications and Signal Processing* (2002)

[95] Mallat, S. "An efficient image representation for multiscale analysis". *Proc. of Machine Vision Conference*, Lake Taho (1997)

[96] Sauer, M. and Götze, J. "Unitary image transforms and their implementation". *Proc. of IEEE Pacific Rim Conference on Communications, Computers and Signal Processing*, 1:55-58 (1993)

[97] Schilling, D. and Cosman, P. "Edge-enhanced image coding for low bit rates". *Proc. of Int. Conf. on Image Processing*, 3:741-751 (1999)

[98] Smith, M. J. and Barnwell, T. P. "A procedure for designing exact reconstruction filter banks for tree structured sub-band coders". *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, San Diego, CA (1984)

[99] Strela, V. and Walden, A. T. "Orthogonal and Biorthogonal Multiwavelets for Signal Denoising and Image Compression," *Proc. SPIE*, 3391:96-107 (1998)

[100] Strömberg, J. O. "A modified Franklin system and higher order spline systems on $\mathfrak{R}^n$ as unconditional bases for Hardy spaces". In Beckner, W., Calderón, A. P., Fefferman, R. and Jones, P. W., eds., *Proc. Conf. in Honor of Antoni Zygmund*, 2:475-493, Wadsworth, New York (1981)

[101] Vass, J., Yao, J., Joshi, A., Palaniappan, K. and Zhuang, X. "Interactive Image Retreival over the Internet". *Proc. of IEEE Symp. on Reliable Distributed Systems*, 1:461-466 (1998)

Research Reports:

[102] Coifman, R. R. and Donoho, D. L. "Translation invariant de-noising". *Tech. Rep.475*, Dept. of Statistics, Standford University (1995)

[103] Gao, H. Y. and Bruce, A. G. "WaveShrink with Semisoft Shrinkage". *Tech. Rep.39*, StatSci Division of MathSoft, Inc. (1995)

[104] Shigeru, F., Yuichiro, N., Se, H. S. and Takefumi, N. "MPEG-4 Video Verification Model version 16.0". *ISO/IEC JTC1/SC29/WG11 Coding of moving pictures and audio, N3312*. International Organisation for Standardization, Noordwijkerhout, pp.371-379 (2000)

[105] "JPEG2000 requirements and profiles version 6.3," *ISO/IEC JTC/SC 29/WG1, Coding for Still Pictures, N1803*. International Organization for Standardization, Jul. (2000)

Web Sites:

[106] T. R. Downie, "Signal Preprocessing of Multiwavelets". http://citeseer.nj.nec.com/cachedpage/117026/1 (1998)