

LOW BIT-RATE COLOR VIDEO COMPRESSION USING MULTIWAVELETS IN
THREE DIMENSIONS

by

Jong-Chih Chien

BSEE, Seattle University, 1990

MSEE, Purdue University, 1993

Submitted to the Graduate Faculty of

School of Engineering in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2005

UNIVERSITY OF PITTSBURGH
SCHOOL OF ENGINEERING

This dissertation was presented

by

Jong-Chih Chien

It was defended on

April 6, 2005

and approved by

John Boston, PhD, Professor, Electrical and Computer Engineering Department

Luis F. Chaparro, PhD, Associate Professor, Electrical and Computer Engineering Department

Heung-no Lee, PhD, Assistant Professor, Electrical and Computer Engineering Department

Shi-Kuo Chang, PhD, Professor, Computer Science Department

Ching-Chung Li, PhD, Professor, Electrical and Computer Engineering Department
Dissertation Director

The author grants to the University of Pittsburgh or its agents the right to archive and display this thesis in whole or in part in all forms of media, now or hereafter known. The author retains all proprietary rights. The author also retains the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

LOW BIT-RATE COLOR VIDEO COMPRESS USING MULTIWAVELETS IN THREE DIMENSIONS

Jong-Chih Chien, PhD

University of Pittsburgh, 2005

In recent years, wavelet-based video compressions have become a major focus of research because of the advantages that it provides. More recently, a growing thrust of studies explored the use of multiple scaling functions and multiple wavelets with desirable properties in various fields, from image de-noising to compression. In term of data compression, multiple scaling functions and wavelets offer a greater flexibility in coefficient quantization at high compression ratio than a comparable single wavelet. The purpose of this research is to investigate the possible improvement of scalable wavelet-based color video compression at low bit-rates by using three-dimensional multiwavelets. The first part of this work included the development of the spatio-temporal decomposition process for multiwavelets and the implementation of an efficient 3-D SPIHT encoder/decoder as a common platform for performance evaluation of two well-known multiwavelet systems against a comparable single wavelet in low bitrate color video compression. The second part involved the development of a motion-compensated 3-D compression codec and a modified SPIHT algorithm designed specifically for this codec by incorporating an advantage in the design of 2D SPIHT into the 3D SPIHT coder. In an experiment that compared their performances, the 3D motion-compensated codec with unmodified 3D SPIHT had gains of 0.3dB to 4.88dB over regular 2D wavelet-based motion-compensated codec using 2D SPIHT in the coding of 19 endoscopy sequences at 1/40 compression ratio. The effectiveness of the modified SPIHT algorithm was verified by the

results of a second experiment in which it was used to re-encode 4 of the 19 sequences with lowest performance gains and improved them by 0.5dB to 1.0dB. The last part of the investigation examined the effect of multiwavelet packet on 3-D video compression as well as the effects of coding multiwavelet packets based on the frequency order and energy content of individual subbands.

TABLE OF CONTENTS

PREFACE.....	xiii
1.0 INTRODUCTION	1
2.0 MULTIWAVELET SYSTEM.....	5
2.1 DEFINITIONS.....	6
2.2 MULTIWAVELET DECOMPOSITION.....	8
2.3 TWO ORTHOGONAL MULTIWAVET SYSTEMS.....	10
2.4 MULTIWAVELET INITIALIZATION.....	14
3.0 COLOR VIDEO COMPRESSION USING 3-D MULTIWAVELETS.....	18
3.1 3-D WAVELET VIDEO COMPRESSION.....	18
3.2 COLOR VIDEO PROCESSING	22
3.3 3-D MULTIWAVELET DECOMPOSITION IN COLOR VIDEO SEQUENCES	24
3.4 3-D SPIHT IN SPATIO-TEMPORAL 3-D WAVELET DOMAIN	30
3.5 EXPERIMENTS AND RESULTS	35
3.6 SUMMARY	42
4.0 3-D MULTIWAVELET-BASED COARSE TRAJECTORY RESIDUALS VIDEO CODING.....	44
4.1 COARSE MOTION RESIDUALS VIDEO CODING.....	45
4.2 KEY FRAME SELECTION STRATEGY FOR 3-D VIDEO CODING.....	52
4.3 TWO ADDITIONAL MULTIWAVELET SYSTEMS.....	54
4.3.1 Four-tapped Symmetric-Antisymmetric Orthogonal Multiwavelet (SA4).....	54
4.3.2 Two-Balanced Cardinal Orthogonal Multiwavelet (Cardbal2)	59
4.4 MODIFIED 3-D SPIHT ENCODER FOR COARSE MOTION RESIDUALS	64

4.5	EXPERIMENTS AND RESULTS	66
4.6	SUMMARY	77
5.0	MULTIWAVELET PACKETS IN 3-D COARSE MOTION RESIDUALS CODING ..	81
5.1	WAVELET PACKETS SPIHT STRUCTURE	86
5.2	PACKET ORDERING BY FREQUENCY	87
5.3	ORDER BY ENERGY	97
5.4	EXPERIMENT AND RESULTS	98
5.5	SUMMARY	106
6.0	CONCLUSIONS AND FUTURE RESEARCH DIRECTIONS.....	107
6.1	SUGGESTIONS FOR FUTURE RESEARCH	109
	APPENDIX.....	111
	RESULTS OF 0.1 BIT-PER-PIXEL VS. 0.25 BIT-PER-PIXEL 3D COMPRESSION	111
	BIBLIOGRAPHY.....	121

LIST OF TABLES

Table 1	Average PSNR over 32 Endoscopy Video Frames at 0.2bpp.....	72
Table 2	Average PSNR over 32 Endoscopy Video Frames at 0.2bpp.....	77
Table 3	Total Residuals summed over 32 Endoscopy Video Frames	80
Table 4	Average PSNR of 3-D Multiwavelet Coarse Motion Residuals Coding with 3D SPIHT Structures (CL).....	99
Table 5	Total Residuals summed over 32 Endoscopy Video Frames Average PSNR of 3-D Multiwavelet Coarse Motion Residuals Coding with 3D SPIHT Structures (SA4).....	100
Table 6	Average PSNR of 3-D Multiwavelet Coarse Motion Residuals Coding with 3D SPIHT Structures (Cardbal2)	100
Table 7	PSNR of 2-D Multiwavelet Coarse Motion Residuals Coding with 2D SPIHT (Chui-Lian) at 0.2bpp.....	103
Table 8	PSNR of 2-D Multiwavelet Coarse Motion Residuals Coding with 2D SPIHT (SA4) at 0.2bpp.....	104
Table 9	PSNR of 2-D Multiwavelet Coarse Motion Residuals Coding with 2D SPIHT (Cardbal2) at 0.2bpp.....	105

LIST OF FIGURES

Figure 1 Geronimo, Hardin and Massopust Multiwavelet.....	13
Figure 2 Chui-Lian Multiwavelet	13
Figure 3 Daubechies 4-coefficients orthogonal wavelet.....	14
Figure 4 Repeated-row Prefiltering for $r=2$	14
Figure 5 Approximation Prefiltering for $r=2$	15
Figure 6 Motion Estimation in the 2-D Wavelet Domain of a Group-of-Pictures	19
Figure 7 Symmetric 3-D Wavelet Decomposition of a Group-of-Pictures	20
Figure 8 The Symmetric 3-D Wavelet Decomposition to the Scale Level 2.....	21
Figure 9 Decoupled 3-D Wavelet Decomposition at Spatial Scale Level 2 and Temporal Scale Level 2	21
Figure 10 (a) A single frame from Mobile video sequence, (b) the same frame with data loss in chrominance U, and (c) the same frame with data loss in chrominance V.....	23
Figure 11 A single frame from the Coastguard video sequence.....	25
Figure 12 Eight sequential luminance frames from Coastguard grouped into a 3-D block	25
Figure 13 A block of 3-D Wavelet Transform (1-level) of the Luminance Portions from 8 Coastguard Frames.....	26
Figure 14 Expanded 3-D Wavelet Blocks of 8 Coastguard Frames in All Three Color Domains	26
Figure 15 One-level Symmetric Three-Dimensional Multiwavelet ($r=2$) Decomposition along 3 Directions.....	27
Figure 16 2-D SPIHT Tree Structure with Dashed Lines indicating Parent-Children Relationships.....	31
Figure 17 Symmetric 3-D SPIHT Tree Structure with Curved Lines indicating Parent-Children Relationships.....	34

Figure 18 Decoupled 3-D SPIHT Tree Structure with curved lines indicating Parent-Children Relationships.....	34
Figure 19 Video frames from (a) Claire, (b) Miss America, (c) Foreman, (d) Silent, (e) Coastguard, (f) Mobile, (g) Salesman, (h) Stefan, and (i) Table Tennis sequences.	36
Figure 20 Experimental Setup of 3D (Multi)Wavelet-based Color Video Compression.....	38
Figure 21 Spatio-Temporal 3-D Video Compression: PSNR Measure of Coastguard Video Sequence at at 0.1 bpp	38
Figure 22 Spatio-Temporal 3-D Compression using CL multiwavelet and DB4 PSNR at 0.33 bpp for luminance portions of (a) Stefan, and (b) Mobile	40
Figure 23 Spatio-Temporal 3-D Compression using CL, GHM Multiwavelets and DB-4, PSNR at 0.33 bpp for Salesman.....	40
Figure 24 Spatio-Temporal 3-D Compression using Multiwavelets (Cl and GHM) and DB4, PSNR at 0.25 bpp for Miss America	42
Figure 25 Object P Travels in Successive Video Frames	44
Figure 26 The Block Diagram of Two-dimensional Wavelet-based Lossy Video Compression	46
Figure 27 Intermediary Model for Lossy Coding of Multiwavelet-based 3-D Video Compression	47
Figure 28 Intermediary Model for Lossy Decoding 3-D Multiwavelet-based Video	47
Figure 29 Threads Describing Motion Trajectories of pixels within a Block in a Video Sequence	49
Figure 30 Intermediary Motion Vectors (Dashed Line) and Coarse Motion Vectors (Solid Line)	50
Figure 31 Scaling Functions (solid) and Wavelets (dashed) of the Symmetric-Antisymmetric Orthogonal Multiwavelet SA4 on the support [0,4]	56
Figure 32 Frequency Responses Plots of SA4 and Chui-Lian Filter Pairs (H-Blue, G-Red).....	58
Figure 33 Two-Balanced Cardinal Multiwavelet Cardbal2 on the support [0,5]; Scaling Functions are shown in Blue/Solid Line, and Wavelets shown in Red/Dashed Line.....	61
Figure 34 Frequency Responses Plots of Cardbal2 Filters	63
Figure 35 Snapshots from Three Neurosurgical Endoscopy Video Sequences.....	67
Figure 36 Search for the Best Match of the NxN Block within a Larger Block.....	68

Figure 37 The Six-loop Full-Search Block Matching Algorithm	69
Figure 38 PSNR Measure of 32 Video Frames of Endoscopy sequence M03	74
Figure 39 PSNR Measure of 32 Video Frames of Endoscopy sequence R02	74
Figure 40 PSNR Measure of 32 Video Frames of Endoscopy sequence R03	75
Figure 41 PSNR Measures of Endoscopy sequence M05	75
Figure 42 PSNR Measures of Endoscopy sequence M08	76
Figure 43 Standard 1-D Multiwavelet ($r=2$) Packet Decomposition for Two Levels	81
Figure 44 Frequency Responses of the Chui-Lian Multiwavelet Filters $\{\tilde{H}_{ij}(w)\}$ and $\{\tilde{G}_{ij}(w)\}$	83
Figure 45 $(\psi_1(k)(x), \psi_2(k)(x)), (k=0,1,2,3)$ of Chui-Lian Multiwavelet Packets at Scale Level 2	84
Figure 46 The Frequency Spectra of the $\psi(k)(\omega) (k=0,1,2,3)$ of the Chui-Lian Multiwavelet ...	85
Figure 47 A Modified “Parent-Children” Relationship for SPIHT in Wavelet Packet Decomposition at Level 2	86
Figure 48 Frequency Coverages of $G_{11}(\omega), H_{11}(\omega)$ and $G_{11}(2\omega), H_{11}(2\omega)$ in the Situation of One Wavelet Basis Function	88
Figure 49 Frequency Ordering of 1D Wavelet Packets for up to Three Levels of Decomposition in the Single Wavelet System	90
Figure 50 Frequency Ordering of 2D Wavelet Packets for up to Two Levels of Decomposition in the Single Wavelet System	91
Figure 51 Frequency Response of $H_{11}(\omega_y)H_{11}(\omega_x)$ for Chui-Lian Multiwavelet.....	92
Figure 52 Frequency Response of $G_{11}(\omega_y)G_{11}(\omega_x)$ for Chui-Lian Multiwavelet.....	93
Figure 53 Frequency Responses of $G_{11}(\omega_y)H_{11}(\omega_x)$ and $H_{11}(\omega_y)G_{11}(\omega_x)$ for Chui-Lian Multiwavelet	93
Figure 54 Frequency Response $H_{11}(\omega_y)H_{11}(2\omega_y)G_{11}(\omega_x)G_{11}(2\omega_x)$ of CL Multiwavelet Filters for LLHH Packet	95
Figure 55 Frequency Response $G_{11}(\omega_y)G_{11}(2\omega_y)G_{11}(\omega_x)G_{11}(2\omega_x)$ of CL Multiwavelet Filters for HHHH Packet	95
Figure 56 The Frequency Spectra of $\psi(k)(\omega) (k=0,1,2,3)$ of the Cardbal2 Multiwavelet.....	96

Figure 57	The Frequency Spectra of the $\psi(k)(\omega)$ ($k=0,1,2,3$) of the SA4 Multiwavelet	97
Figure 58	CL, GHM multiwavelets, and DB4 PSNR at 0.1 bpp and 0.25 bpp for Claire.....	112
Figure 59	CL, GHM multiwavelets, and DB4 PSNR at 0.1 bpp and 0.25 bpp for Coastguard	113
Figure 60	CL, GHM multiwavelets, and DB4 PSNR at 0.1 bpp and 0.25 bpp for Foreman	114
Figure 61	CL, GHM multiwavelets, and DB4 PSNR at 0.1 bpp and 0.25 bpp for Miss America	115
Figure 62	CL, GHM multiwavelets, and DB4 PSNR at 0.1 bpp and 0.25 bpp for Mobile.....	116
Figure 63	CL, GHM multiwavelets, and DB4 PSNR at 0.1 bpp and 0.25 bpp for Salesman ...	117
Figure 64	CL, GHM multiwavelets, and DB4 PSNR at 0.1 bpp and 0.25 bpp for Silent	118
Figure 65	CL, GHM multiwavelets, and DB4 PSNR at 0.1 bpp and 0.25 bpp for Stefan	119
Figure 66	CL, GHM, and DB4 PSNR at 0.1 bpp and 0.25 bpp for Table Tennis.....	120

PREFACE

I would like to like my committee members: Dr. Boston, Dr. Chaparro, Dr. Lee, Dr. Chang from the Department of Computer Science and especially my advisor Dr. Li for their support over the years.

I would also like to thank my family in Taiwan and my sister and brother-in-law in the US for their love and support.

Lastly, by the Grace of Our Lord, I would very much like to thank all my brothers and sisters in my church for their prayers, unwavering support and faith despite the difficulties.

1.0 INTRODUCTION

The paradigm of wavelet-based image compression may be like viewing of a museum masterpiece; first you view it from afar to get the impact of the whole picture in coarse resolution, then additional pieces of finer information are added as you step closer until you have completed viewing of the entire masterpiece. The spatio-temporal 3-dimensional wavelet-based video compression is a natural extension to this paradigm, where the masterpieces were all drawn by the same master and the shared information in these masterpieces can be first viewed as a coarse 3-dimensional outline, and then the differences can be added as we view them in greater detail. The main objective of this dissertation research is to explore the application of the spatio-temporal three-dimensional multiwavelets to color video compression as a good model that fits this paradigm. The motivation behind this objective is to improve the performance of low bit-rate scalable color video compression by using three-dimensional multiwavelet transform instead of the single wavelet transform

There has been a growing number of studies on applications of multiwavelets in various fields, from image denoising[9,11-17] and image [2-10,25,26] and video compression[75], to solving nonlinear equations [21] where representations with multiple sets of basis functions may offer some advantages, are just a few among many studies[1,18-20,22-24,27-29,30-32,34]. However, little has been done on applying spatio-temporal 3-D multiwavelets to video compression, although studies have been reported on using 3-D single wavelets on using different decomposition structures [38-44,46], on coding improvements [45,47,48,51,71,72], and on codec

designs for specific requirements[49,50,52-56,58,59]. The one obvious advantage revealed in all these studies is that three-dimensional wavelet-based video coding is that it can provide easier scalability in both spatial and temporal domains (frame size and rate change) at the same time than two dimensional video coding. Also, Xu, et. al[75] mentioned, in his study of a motion compensated three-dimensional single wavelet coder, that greater gains may be achieved by incorporating motion information into the 3-D video coder.

In 3-D wavelet-based video coding, a tensor product of two-dimensional wavelet on transform in the spatial domain and a one-dimensional wavelet transform in the temporal domain is performed on a sequence of video frames. To perform the transform along the temporal dimension, a number of frames called group-of-pictures (GOP) must be buffered first, and in order to minimize processing delay, the temporal span of GOP is relatively small in comparison to the spatial span of the frames. The three-dimensional wavelet transform using a single wavelet has been shown to be efficient for video [39,48-51,71] and medical (MRI and CT) image sequence compression [52-54,58], while merits of the two-dimensional spatial multiwavelet transform has been investigated in image and video compression, classification, and denoising [1-10,12,13,15-18,20,22,25,26,71]. Multiwavelets are designed to combine several desirable properties together in the matrix filters which would not be possible for the single wavelet filters; these properties include orthogonality, symmetry, high approximation order and compact support. These properties together enable multiwavelet to better represent the signal using fewer coefficients than a comparable single wavelet. In addition, the scaling and wavelet coefficients obtained using multiple scaling functions and multiple wavelets can offer greater flexibility during the process of coefficient quantization. In this dissertation, we will show that

mulwavelets, with the advantages mentioned above, will give better results on 3-D color video compression at higher compression ratios than those obtainable by single wavelet.

The studies on multiwavelet-based compression of natural images[4,9,26,75] have shown some improvements, especially at high compression ratios, in comparison to the compression results obtained using single wavelet with comparable properties, although these improvements were not extraordinary. Bell et.al [2] compared color image compression using multiwavelet against that using single single wavelet, and showed that certain multiwavelets produce slightly better results both subjectively and qualitatively. However, the studies on signal and image denoising by thresholding wavelet coefficients have shown very interesting results: Strela et.al [9] found that some multiwavelets, such as the Chui-Lian multiwavelet, demonstrated obvious advantages over single wavelets under very noisy conditions. Similar conditions exist in the low bit-rate color videos. Since the effect of wavelet shrinkage used in denoising is similar to that of bit plane-based encoding of wavelet coefficients, where only the most significant bits of data are preserved under high compression, it can be expected that the same multiwavelets that achieved good results in image denoising can also be used to improve the three-dimensional wavelet-based video coding in low band-width environments. Therefore, this is the motivation behind this thesis in the study of the advantages of spatio-temporal three-dimensional multiwavelets in exploiting the redundancy in color videos for lossy compression. Tham [75] had proposed a framework for multiwavelet-based two-dimensional video compression with motion estimation in transform domain. His simulation results included an average PSNR of 23.77dB at 0.25 bit-per-pixel, and an average PSNR 25.06dB on a 0.79 bit-per-pixel compression ratio for the QCIF-size grayscale Carphone sequence. We believe his multiwavelet-based model did not take full advantage of the characteristics of multiwavelets, and of the reasons we believe this is the well-

known problem of shift-variance[34-37] which could lead to sub-optimal results when motion estimation is performed in the wavelet domain and reduce compression performance. We also believe that a two-dimensional video model could not fully exploit the advantages of multiple scaling functions and multiwavelet wavelets as well as three-dimensional model. So despite Tham's work in multiwavelet-based video compression, there is still a question of performance comparison between multiwavelet and single wavelet in terms of 3D video compression which this dissertation will answer.

This thesis is organized as follows. First, a brief review of multiwavelets and related background is given in Chapter 2. Chapter 3 presents the work on three-dimensional multiwavelet-based video coding. Chapter 4 presents the study of three-dimensional multiwavelet-based coarse-motion-trajectory residuals video coding and compares experimentally its performance to that of the standard two-dimensional single-wavelet and multiwavelet motion-trajectory-residuals video coding. In Chapter 5, multiwavelet packets and the shuffling of subbands are considered in three-dimensional video coding. Chapter 6 summarizes the major contributions of this research and suggests topics for future research.

2.0 MULTIWAVELET SYSTEM

Multiwavelets as basis functions possess several desirable properties which are not possible to have simultaneously in a single wavelet system, such as orthogonality, symmetry/antisymmetry, compact support and high vanishing moment/approximate order. These properties are explained in the following:

- Orthogonality

Signals can be efficiently represented in terms of orthogonal basis functions that introduce no redundant information and are also energy preserving, they give support to efficient data analysis.

- Symmetry/Antisymmetry

Symmetry of the basis functions leads to symmetry of filters giving the property of linear phase. This is important in data encoders, such as SPIHT, that use a predefined hierarchical structure incorporating correlations between the data in the children nodes and those in the parent node, so that the data in the parent node can be used to predict the data in the children nodes.

- High Approximation Order/Vanishing Moment

The smoothness of a wavelet representation is commensurate with the highest order at which the wavelet's moment vanishes. Wavelet representation with high degree of

smoothness minimizes distortions in the approximation and will likely lead to a representation with only a few coefficients.

- Compact Support

Compactly supported scaling functions and the associated wavelets lead to filters whose coefficients are zeros outside of a finite support range. So, a filter with a shorter support is computationally less expensive than a filter with a longer support.

In this chapter we will briefly review the concept of multiwavelets and the associated background

2.1 DEFINITIONS

Let us consider the multiresolution analysis (MRA) in $L^2(\mathbb{R})$ given by a sequence of scaling subspaces $\{V_j\}_{j \in \mathbb{Z}}$,

$$\{0\} \subset \dots \subset V_{-1} \subset V_0 \subset V_1 \subset \dots \subset L^2(\mathbb{R})$$

where the subscript j denotes the scale index, and larger j means finer scale. V_0 is generated by a scaling function $\phi(x)$, whose integer translates are orthogonal, and V_j is spanned by $\{\phi_{j,k}(x)\}$, where

$$\phi_{j,k}(x) := 2^{j/2} \phi(2^j x - k), \quad \forall x \in \mathbb{R}$$

Corresponding to $\phi(x)$ there is a wavelet $\psi(x)$. $\phi(x)$ and $\psi(x)$ satisfy the 2-scale dilation equations:

$$\phi(x) = 2 \sum_n^{N_1} h_n \phi(2x - n)$$

$$\psi(x) = 2 \sum_n^{N_2} g_n \phi(2x - n)$$

Let us consider a subspace W_j spanned by $\{\psi_{j,k}(x)\}$ where

$$\psi_{j,k}(x) := 2^{j/2} \psi(2^j x - k), \quad \forall x \in \mathbb{R}$$

such that $V_{j+1} = V_j \oplus W_j$, where \oplus denotes the direct sum. So W_j complements V_j in V_{j+1} and

$$W_j \cap W_l = \{0\} \quad \forall j \neq l$$

One may find a set of multiple scaling functions $\phi^1, \phi^2, \dots, \phi^r$ and their associated multiple wavelets $\psi^1, \psi^2, \dots, \psi^r$, where r denotes the multiplicity of the multiwavelets, such that V_j is a closure of the linear span of $\{\phi_{j,k}^m\}_{1 \leq m \leq r, k \in \mathbb{Z}}$ in $L^2(\mathbb{R})$,

$$\phi_{j,k}^m(x) := 2^{j/2} \phi^m(2^j x - k), \quad \forall x \in \mathbb{R}$$

and W_j is a closure of the linear span of $\{\psi_{j,k}^m\}_{1 \leq m \leq r, k \in \mathbb{Z}}$ in $L^2(\mathbb{R})$,

$$\psi_{j,k}^m(x) := 2^{j/2} \psi^m(2^j x - k), \quad \forall x \in \mathbb{R}$$

Now let the multi-scaling functions and associated multiwavelets be expressed in the form of vectors:

$$\phi(x) := \begin{pmatrix} \phi^1(x) \\ \phi^2(x) \\ \vdots \\ \phi^r(x) \end{pmatrix} \quad \forall x \in \mathbb{R}$$

and

$$\psi(x) := \begin{pmatrix} \psi^1(x) \\ \psi^2(x) \\ \vdots \\ \psi^r(x) \end{pmatrix} \quad \forall x \in \mathbb{R}$$

$$\phi_{j,k}(x) := 2^{j/2} \phi(2^j x - k), \quad \forall x \in \mathbb{R}$$

$$\psi_{j,k}(x) := 2^{j/2} \psi(2^j x - k), \quad \forall x \in \mathbb{R}$$

Because $\phi^m \in V_0$, $\psi^m \in W_0$, $1 \leq m \leq r$, and $V_0, W_0 \subset V_1$ is generated by

$$\{\phi_{1,k}^m(x) = 2^{j/2}\phi^m(2x-k)\}_{1 \leq m \leq r, k \in \mathbf{Z}},$$

there exist two matrix sequences $\{H_n\}_{n \in \mathbf{Z}}$ and $\{G_n\}_{n \in \mathbf{Z}}$ such that the matrix two-scale dilation equations holds for the multi-scaling functions Φ and multiwavelets Ψ :

$$\phi(x) = 2 \sum_{n=0}^N H_n \phi(2x-n), \quad \forall x \in \mathbb{R}$$

$$\psi(x) = 2 \sum_{n=0}^N G_n \psi(2x-n), \quad \forall x \in \mathbb{R}$$

where H_n and G_n are $r \times r$ matrices. The support of the multiscaling functions $\phi(x)$ is defined as

$$\text{support}(\phi) = \bigcup \text{support}(\phi^m), \quad 1 \leq m \leq r,$$

which is bounded by $[0, N]$, then n ranges from 0 to N .

$H = \{H_n\}$ is the low-pass multifilter, and $G = \{G_n\}$ is the high-pass multifilter used in the analysis.

Since $\phi(2x-k)$ are in V_1 and $V_1 = V_0 \oplus W_0$, there exists matrix sequences $\{\tilde{H}_n\}_{n \in \mathbf{Z}}$ and $\{\tilde{G}_n\}_{n \in \mathbf{Z}}$ such that

$$\phi(2x-k) = \sum_{n \in \mathbf{Z}} [\tilde{H}_{k-2n}^T \phi(x-n) + \tilde{G}_{k-2n}^T \psi(x-n)], \quad \forall k \in \mathbf{Z}$$

which gives the synthesis relation between $\phi(2x)$ and $\phi(x)$ and $\psi(x)$ [25]; in the orthogonal multiwavelet system, $\tilde{H}_n = H_n$ and $\tilde{G}_n = G_n$.

2.2 MULTIWAVELET DECOMPOSITION

The multiwavelet decomposition and reconstruction of a 1-D signal can be derived from the two-scale equations and the decomposition relationship, just as in the single wavelet case, where $\{c_{j-1,k}\}_{j,k \in \mathbf{Z}}$ and $\{d_{j-1,k}\}_{j,k \in \mathbf{Z}}$ are two coefficient vectors sequences, respectively. Let $v_{j-1}(x) \in V_{j-1}$ and $w_{j-1}(x) \in W_{j-1}$ be the scaling component and wavelet component of the signal at the scale level $(j-1)$ respectively,

$$v_{j-1}(x) := \sum_{n \in \mathbb{Z}} c_{j-1,n}^T \phi(2^{j-1}x - n) \quad \text{and} \quad w_{j-1}(x) := \sum_{n \in \mathbb{Z}} d_{j-1,n}^T \psi(2^{j-1}x - n)$$

For the sake of simplicity, the $2^{j/2}$ factor is incorporated into the coefficient sequences rather than explicitly shown. By the relationship $V_j := V_{j-1} \oplus W_{j-1}$, we have

$$v_j(x) = v_{j-1}(x) + w_{j-1}(x) = \sum_{n \in \mathbb{Z}} c_{j-1,n}^T \phi(2^{j-1}x - n) + \sum_{n \in \mathbb{Z}} d_{j-1,n}^T \psi(2^{j-1}x - n)$$

However, by using the synthesis relation, we have:

$$\begin{aligned} v_j(x) &= \sum_{k \in \mathbb{Z}} c_{j,k}^T \phi(2^j x - k) \\ &= \sum_{k \in \mathbb{Z}} c_{j,k}^T \left(\sum_{n \in \mathbb{Z}} [\hat{H}_{k-2n}^T \phi(2^{j-1}x - n) + \tilde{G}_{k-2n}^T \psi(2^{j-1}x - n)] \right) \\ &= \sum_{n \in \mathbb{Z}} \left[\sum_{k \in \mathbb{Z}} c_{j,k}^T \hat{H}_{k-2n}^T \right] (\phi(2^{j-1}x - n)) + \sum_{n \in \mathbb{Z}} \left[\sum_{k \in \mathbb{Z}} d_{j,k}^T \tilde{G}_{k-2n}^T \right] (\psi(2^{j-1}x - n)) \end{aligned}$$

Thus, we have the recursive decomposition formulas for 1-D signals:

$$\begin{aligned} c_{j-1,n} &= \sum_k \tilde{H}_{k-2n} c_{j,k} = \sum_k \tilde{H}_{-k} c_{j,2n-k} \\ d_{j-1,n} &= \sum_k \tilde{G}_{k-2n} d_{j,k} = \sum_k \tilde{G}_{-k} c_{j,2n-k} \end{aligned}$$

If an original input vector sequence is c_0 , we can decompose it into c_{-1} and d_{-1} sequences, and if so desired, decomposed c_{-1} to further coarser scale levels.

To reconstruct the signal from its decomposition, we have

$$\begin{aligned} v_j(x) &= v_{j-1}(x) + w_{j-1}(x) = \sum_{n \in \mathbb{Z}} c_{j-1,n}^T \phi(2^{j-1}x - n) + \sum_{n \in \mathbb{Z}} d_{j-1,n}^T \psi(2^{j-1}x - n) \\ &= \sum_{n \in \mathbb{Z}} c_{j-1,n}^T 2 \sum_m H_m \phi(2^j x - 2n - m) + \sum_{n \in \mathbb{Z}} d_{j-1,n}^T 2 \sum_m G_m \phi(2^j x - 2n - m) \end{aligned}$$

Let $k=2n+m$, then

$$v_j(x) = \sum_k 2 \left[\sum_n c_{j-1,n}^T H_{k-2n} + \sum_n d_{j-1,n}^T G_{k-2n} \right] \phi(2^j x - k) = \sum_k c_{j,k}^T \phi(2^j x - k)$$

Hence, we have the recursive formula for reconstruction:

$$c_{j,k} = 2 \sum_n (H_{k-2n}^T c_{j-1,n} + G_{k-2n}^T d_{j-1,n})$$

For biorthogonal multiwavelet system, the decomposition and reconstruction filters must satisfy the following equations:

$$(i) \quad 2 \sum_{n \in \mathbb{Z}} H_n \hat{H}_{n-2k}^T = \delta I_r$$

$$(ii) \quad 2 \sum_{n \in \mathbb{Z}} H_n \hat{H}_{n-2k}^T = 0$$

$$(iii) \quad 2 \sum_{n \in \mathbb{Z}} H_n \tilde{G}_{n-2k}^T = 0, \text{ and}$$

$$(iv) \quad 2 \sum_{n \in \mathbb{Z}} G_n \tilde{G}_{n-2k}^T = \delta I_r$$

In terms of the Z-transform of the filter matrices,

$$H(Z) \tilde{H}(Z) + H(-Z) \tilde{H}^*(-Z) = I_r$$

$$G(Z) \hat{H}^*(Z) + G(-Z) \hat{H}^*(Z) = 0$$

$$H(Z) \tilde{G}(Z) + H(-Z) \tilde{G}^*(-Z) = I_r$$

$$G(Z) \tilde{G}^*(Z) + G(-Z) \tilde{G}^*(-Z) = I_r$$

In case of orthogonal multiwavelet systems, $\tilde{H}(Z) = H(Z)$, $\tilde{H}_n = H_n$, $\tilde{G}(Z) = G(Z)$ and $\tilde{G}_n = G_n$

2.3 TWO ORTHOGONAL MULTIWAVET SYSTEMS

Two well-known orthogonal multiwavelet systems (for $r=2$) are the Geronimo-Hardin-Massopust[32] and the Chui-Lian[33] families of multiwavelets. The filter coefficients of $\{H_n, G_n\}$ of the Geronimo-Hardin-Massopust (GHM) multiwavelet are

$$H_0 = \begin{bmatrix} \frac{3}{10} & \frac{4}{10\sqrt{2}} \\ \frac{1}{20\sqrt{2}} & -\frac{3}{20} \end{bmatrix}, \quad H_1 = \begin{bmatrix} \frac{3}{10} & 0 \\ \frac{9}{20\sqrt{2}} & \frac{1}{2} \end{bmatrix}, \quad H_2 = \begin{bmatrix} 0 & 0 \\ \frac{9}{20\sqrt{2}} & -\frac{3}{20} \end{bmatrix}, \quad H_3 = \begin{bmatrix} 0 & 0 \\ -\frac{1}{20\sqrt{2}} & 0 \end{bmatrix}$$

and

$$G_0 = \begin{bmatrix} -\frac{1}{20\sqrt{2}} & -\frac{3}{20} \\ \frac{1}{20} & \frac{3}{20\sqrt{2}} \end{bmatrix}, \quad G_1 = \begin{bmatrix} \frac{9}{20\sqrt{2}} & -\frac{1}{2} \\ -\frac{9}{20} & 0 \end{bmatrix}, \quad G_2 = \begin{bmatrix} \frac{9}{20\sqrt{2}} & -\frac{3}{20} \\ \frac{9}{20} & \frac{3}{10\sqrt{2}} \end{bmatrix}, \quad G_3 = \begin{bmatrix} -\frac{1}{20\sqrt{2}} & 0 \\ -\frac{1}{20} & 0 \end{bmatrix}$$

The Geronimo-Hardin-Massopust multiwavelet possesses the following properties:

- The scaling functions $\phi^1(x)$ and $\phi^2(x)$ have very short supports of $[0,1]$ and $[0,2]$, respectively.
- Both scaling functions are symmetric while the wavelet functions form a symmetric/antisymmetric pair as shown in Figure 1.
- They are orthonormal.
- They provide the second order approximation.

The filter coefficients $\{H_n, G_n\}$ of the Chui-Lian (CL) multiwavelet are

$$H_0 = \begin{bmatrix} \frac{1}{4} & \frac{1}{4} \\ -\frac{\sqrt{7}}{8} & -\frac{\sqrt{7}}{8} \end{bmatrix}, \quad H_1 = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{4} \end{bmatrix}, \quad H_2 = \begin{bmatrix} \frac{1}{4} & -\frac{1}{4} \\ -\frac{\sqrt{7}}{8} & -\frac{\sqrt{7}}{8} \end{bmatrix}$$

and

$$G_0 = \begin{bmatrix} -\frac{1}{4} & -\frac{1}{4} \\ \frac{1}{8} & \frac{1}{8} \end{bmatrix}, \quad G_1 = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{\sqrt{7}}{4} \end{bmatrix}, \quad G_2 = \begin{bmatrix} -\frac{1}{4} & \frac{1}{4} \\ -\frac{1}{8} & \frac{1}{8} \end{bmatrix}$$

The Chui-Lian multiwavelet possesses the following properties:

- Both scaling functions have very short support of $[0,2]$.

- Both scaling functions and wavelet functions are symmetric/antisymmetric pairs, as shown in Figure 2.
- They are orthonormal.
- They provide the second order approximation.

For the purpose of comparison, Figure 3 shows the scaling function ϕ and wavelet ψ of the 4-tapped Daubechies DB4 orthonormal wavelet system, both are supported in $[0,4]$ and have filter coefficients of length 4 as given by:

$$\{h_n\} = \left[\frac{1+\sqrt{3}}{8} \quad \frac{3+\sqrt{3}}{8} \quad \frac{3-\sqrt{3}}{8} \quad \frac{1-\sqrt{3}}{8} \right]$$

and

$$\{g_n\} = \left[\frac{1-\sqrt{3}}{8} \quad \frac{-3+\sqrt{3}}{8} \quad \frac{3+\sqrt{3}}{8} \quad \frac{-1-\sqrt{3}}{8} \right]$$

It also has the second order approximation property, but the scaling function of DB4 is neither symmetric nor antisymmetric. As noted earlier, the single wavelet system can not possess all these desirable properties simultaneously.

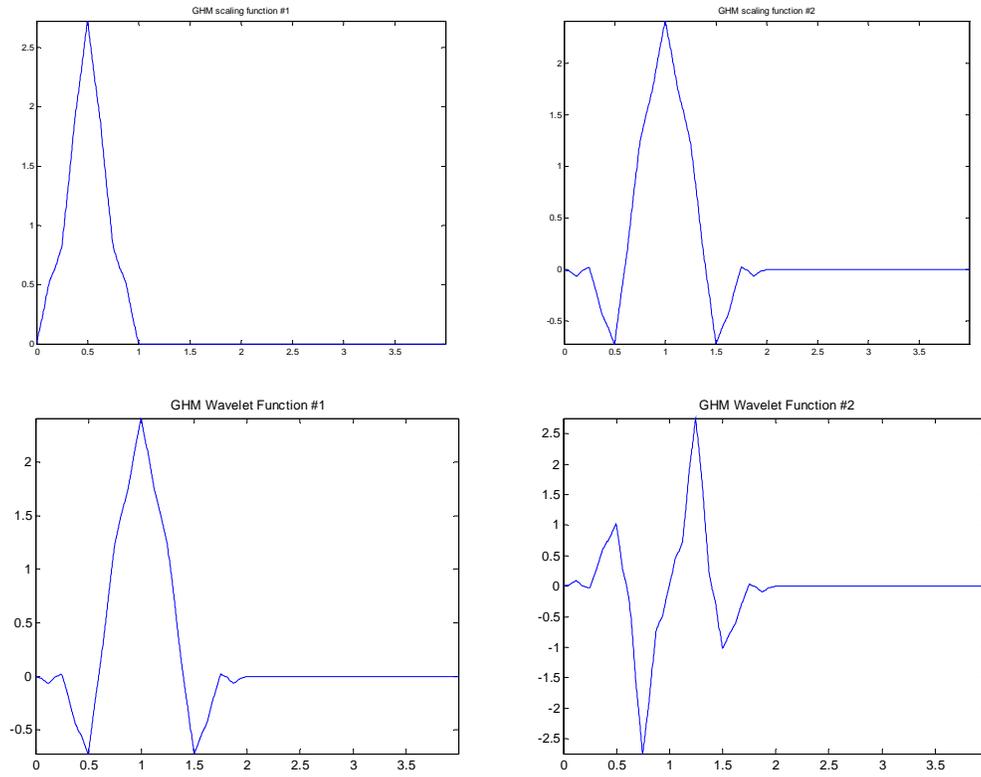


Figure 1 Geronimo, Hardin and Massopust Multiwavelet

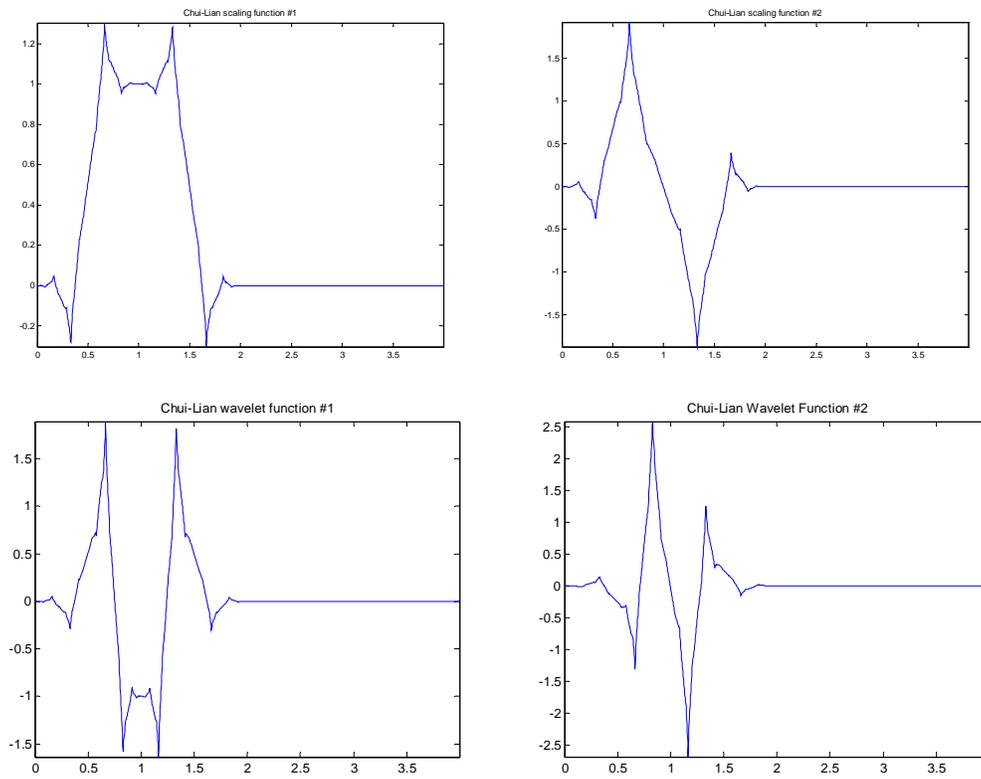


Figure 2 Chui-Lian Multiwavelet

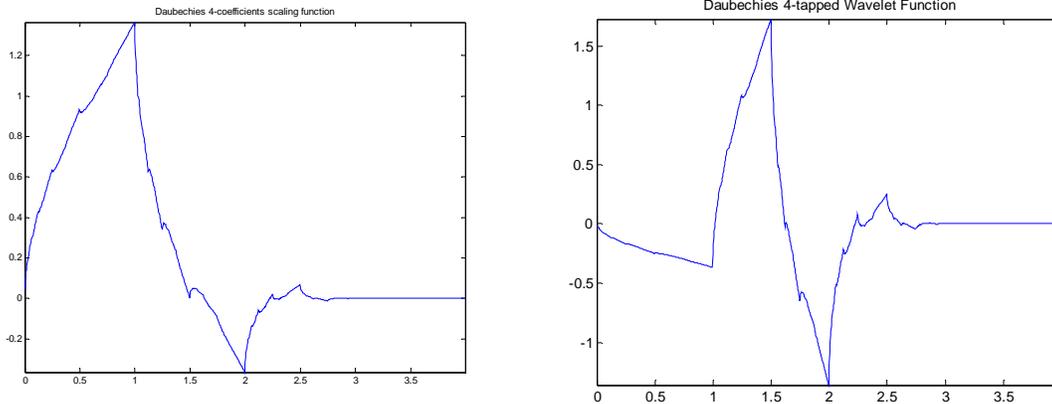


Figure 3 Daubechies 4-coefficients orthogonal wavelet

2.4 MULTIWAVELET INITIALIZATION

When using multiwavelet basis functions for signal decomposition, one must deal with the problem of how to adapt a discrete-time signal to matrix filters [23-30]. We will review briefly some methods of preconditioning an input signal into a vector format for the matrix filters. A 1-D signal must be converted into two input rows for the multiwavelet system ($r=2$). This is called the preconditioning or initialization problem. One obvious way is to send the signal to each row (channel) as shown in Figure 4; this would be useful for feature extraction in data analysis, but would not be useful for data compression as it introduces additional redundancy.

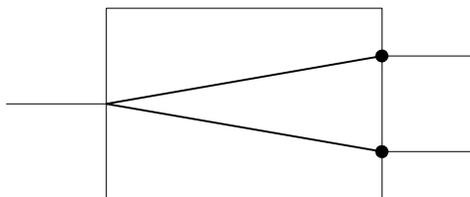


Figure 4 Repeated-row Prefiltering for $r=2$

We would like to generate multiple (2) rows of data from the input by sampling, and distribute the original data amongst the rows, so the resultant number of data points is the same as in the original input. This is termed the critically-sampled scheme. The idea is to find a set of

preprocessing filters $Q(z)$ that will sample and process the input to provide a vector-format data as the input for multiwavelet decomposition, and a set of post processing filters $P(z)$ such that $P(z)Q(z)=z^{-\ell}I$, $\ell \in \mathbf{Z}$, to guarantee the perfect reconstruction. Two critically-sampled schemes are described below.

Strela [9] introduced this scheme based on the approximation properties of the continuous-time multiwavelets, and has applied it successfully in 2-D compression of synthetic images and denoising applications. Figure 5 illustrate the block layout of the approximation-based preprocessing for the case of $r=2$. The problem of obtaining approximation-based preprocessing filters is formulated as follows:

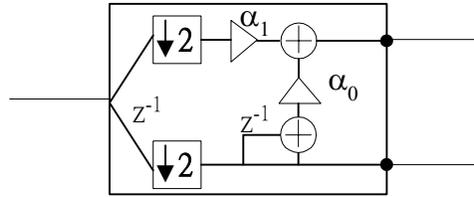


Figure 5 Approximation Prefiltering for $r=2$

Let the continuous-time function $f(t) \in V_0$ be generated by translates of the scaling functions so that it is formulated by two sets of linear combination of the translates.

$$f(t) = \sum_n q_{1,n} \phi^1(t-n) + \sum_n q_{2,n} \phi^2(t-n)$$

To obtain $q_{1,n}$ and $q_{2,n}$, let us suppose that the input $F[n]$ contains samples of $f(t)$ at half-integers so that $F[2n]=f(n)$ and $F[2n+1]=f(n+\frac{1}{2})$. For a multiwavelet such as GHM family where $\phi^1(t)$ vanishes at all integer points, and $\phi^2(t)$ also vanishes at integer points except at integer 1, then sampling $f(t)$ at integers and half-integers gives

$$F[2n]=\phi^2(1)q_{2,n-1}, G[2n+1]=\phi^2(\frac{3}{2})q_{2,n-1}+\phi^1(\frac{1}{2})q_{1,n}+\phi^2(\frac{1}{2})q_{2,n}$$

then

$$q_{1,n} = \frac{\phi^2(1)F[2n+1] - \phi^2(\frac{1}{2})(F[2n+2] + \phi^2(\frac{3}{2})F[2n])}{\phi^2(1)\phi^1(\frac{1}{2})}, \text{ and } q_{2,n} = \frac{F[2n+2]}{\phi^2(1)}$$

If $\phi^2(t)$ is furthermore symmetric, then

$$q_{1,n} = \frac{\phi^2(1)F[2n+1] - \phi^2(\frac{1}{2})(F[2n+2] + F[2n])}{\phi^2(1)\phi^1(\frac{1}{2})}, \text{ and } q_{2,n} = \frac{F[2n+2]}{\phi^2(1)}$$

which naturally yields a critically sampled, preprocessing to generate two more balanced rows of data from a single row. Note that the above algorithm for $r=2$ can be easily generalized to $r \geq 2$.

Similar to the approximation preprocessing described above, the interpolation-based preprocessing introduced by Kim and Li[25,26] seeks to reduce undesirable artifacts introduced in data subsampling by formulating the problem of prefiltering as an interpolation problem:

Let a continuous signal $f_0(t) := \sum_k c_{0,k}^T \phi(t-k)$ interpolate the original discrete signal $f[n]$. For r

scaling functions, $f_0(t)$ is sampled at the interval of $\frac{1}{r}$, so that

$$f_0\left(\frac{n}{r}\right) = \sum_{k \in \mathbb{Z}} c_{0,k}^T \phi\left(\frac{n}{r} - k\right) = \sum_{k \in \mathbb{Z}} \phi^T\left(\frac{n}{r} - k\right) c_{0,k}$$

with the condition that $f_0\left(\frac{n}{r}\right) = f[n]$. Let vectors $\mathbf{f}_0[n]$ and $\mathbf{f}[n]$ be constructed from $f_0\left(\frac{n}{r}\right)$ and $f[n]$,

respectively, such that

$$\mathbf{f}_0[n] := \begin{bmatrix} f_0(n) \\ f_0\left(n + \frac{1}{r}\right) \\ \vdots \\ f_0\left(n + \frac{r-1}{r}\right) \end{bmatrix} \quad \text{and} \quad \mathbf{f}[n] := \begin{bmatrix} f[nr] \\ f[nr+1] \\ \vdots \\ f[nr+r-1] \end{bmatrix}$$

By applying the condition $f_0(\frac{n}{r})=f[n]$, we have

$$\mathbf{f}_0[n] = \mathbf{f}[n] = \left[\dots \phi^T\left(\frac{nr+m}{r} * k\right) \dots \right] \begin{bmatrix} \vdots \\ c_{0k} \\ \vdots \end{bmatrix} = \sum_{k \in \mathbb{Z}} P_{n-k} c_0[k] = \sum_{k \in \mathbb{Z}} P_k c_0[n-k],$$

$$\text{where } P_n := \begin{bmatrix} \phi^T(n) \\ \phi^T(n+\frac{1}{r}) \\ \vdots \\ \phi^T(n+\frac{r-1}{r}) \end{bmatrix}$$

which is the interpolation postfilter that maps multiscaling coefficients $c_0[n-k]$ back to the space of sampled signals. Then the prefilter $Q(z)$ can be obtained via $P(z)Q(z)=z^{-\ell}I=I_r$, by letting $\ell=0$. It should be noted that stable interpolation prefilters can not be obtained for all kinds of multiwavelets.

3.0 COLOR VIDEO COMPRESSION USING 3-D MULTIWAVELETS

In this chapter, we introduce the application of spatio-temporal three-dimensional multiwavelet transform to color video compression by developing a multi-wavelet based 3-D compression model. This model is then tested on a set of color video sequences at various low compression ratios.

3.1 3-D WAVELET VIDEO COMPRESSION

Wavelet-based compression is similar, in one part of the basic operations, to wavelet-based denoising. In wavelet-based denoising, a noisy image is first transformed into the wavelet domain with scaling coefficients and wavelet coefficients. Based on the statistics of the coefficients in each wavelet subband, a threshold is computed for each subband. The thresholded wavelet coefficients together with the original scaling coefficients are then used in the inverse transform to reconstruct an image that approximates the original image but with significant reduction in noise. This process is termed wavelet shrinkage. There are two different ways to perform thresholding: hard threshold and soft threshold. In hard thresholding, wavelet coefficients that are above the threshold value remain unchanged, while those coefficients below the threshold are set to zero. Strela [9] performed hard thresholding on multiwavelet coefficients and obtained better denoised results when measured against those obtained by using single wavelet. His results were especially striking when the Chui-Lian multiwavelet was used on synthetic images. In soft thresholding, wavelet coefficients above the threshold are altered by

reducing the coefficient values by the amount of the threshold. In this way, the soft thresholding avoids drastic discontinuities in wavelet coefficients used in reconstruction of a denoised image, thus avoiding undesirable artifacts introduced by the discontinuities. Recent studies on multiwavelets shrinkage using soft thresholding [14,16] showed that the operation approximates a bit-plane based compression of wavelet coefficients, thus the results obtained by Strela [9] on noisy images may be translated to 3-D video compression.

In video coding, high levels of correlation naturally exist between video frames generated from the same source. One method to remove the redundancy between video frames is to applying the 2-D discrete wavelet transform to each frame, compute motion prediction across the frames to remove the temporal redundancy, and finally encode the prediction errors as illustrated in Figure 6. [48,51,52] This method can yield high level of compression but suffers from some inherent problems:

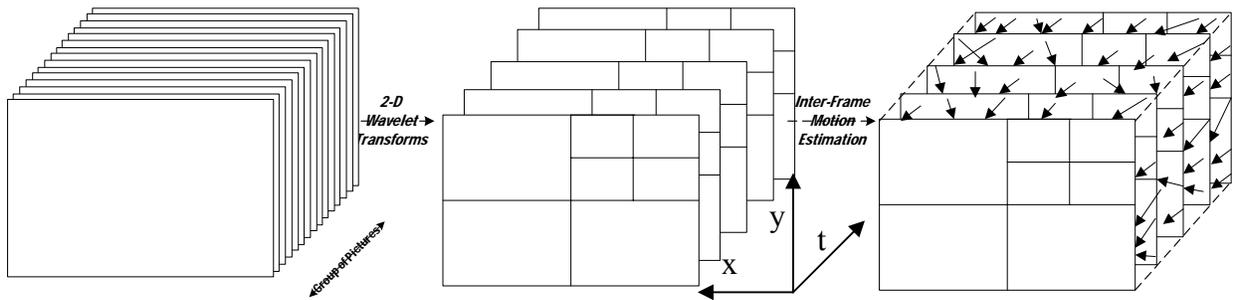


Figure 6 Motion Estimation in the 2-D Wavelet Domain of a Group-of-Pictures

1. Wavelet coefficients describing the same moving object may not be invariant from frame to frame, thus causing inaccuracy in motion estimation. [34,35]
2. Block-based motion estimation often produces blocky artifacts in the video and is computationally expensive [35].

3. Since the quality of each decoded frame is dependent on the quality of its reference frame, any corruption in the data will be propagated to successive frames and get enlarged.

The second method is to partition the video into successive segments of video frames, each segment is considered as a spatio-temporal 3-D data block and the 3-D wavelet transform is applied to the data block as illustrated in Figure 7, so redundancies in both spatial and temporal domains are projected into the scaling space and wavelet space making it easier to be prioritized for removal. The standard 3-D wavelet decomposition can be computed by the 3-dimensional tensor product as a direct extension of the 2-D discrete wavelet transform by tensor product. Note that the tensor product of two 1-D discrete wavelet transforms must be at the same scale level. This is called the symmetric decomposition as employed in Figure 8 in 3-D.

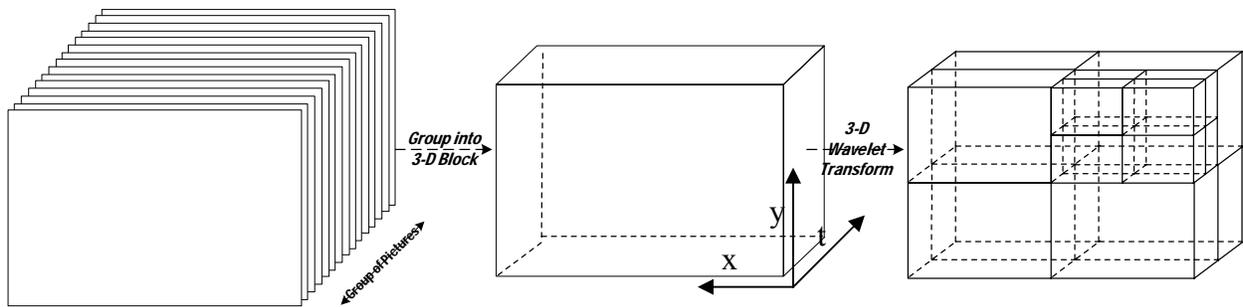


Figure 7 Symmetric 3-D Wavelet Decomposition of a Group-of-Pictures

Since the data supports along the time axis and along the spatial axes may be very different, one may wish to decompose to different levels, along the temporal direction and along the spatial directions. Hence, another way of computing the 3-D discrete wavelet decomposition is to perform decomposition along the spatial directions to one scale level, say J , and then continue the composition along the temporal direction to a scale level, Q , which may be different from J . This is illustrated in Figure 9. This decomposition, which is called the decoupled

decomposition, may give wavelet coefficients which are partly regular wavelet coefficients and partly wavelet packet coefficients – a different data structure than the symmetric decomposition.

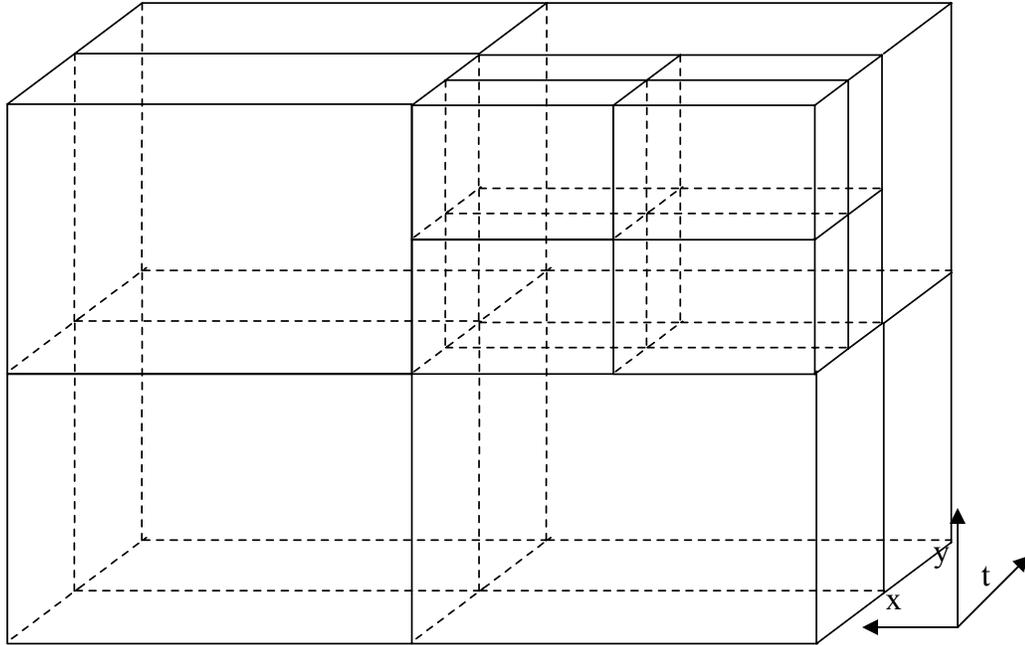


Figure 8 The Symmetric 3-D Wavelet Decomposition to the Scale Level 2

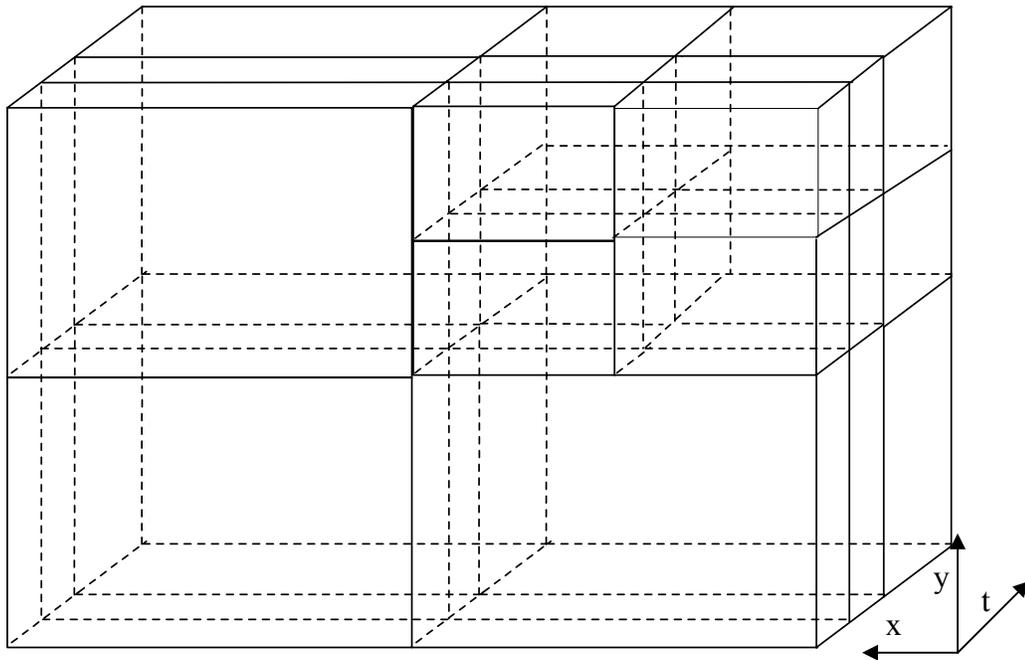


Figure 9 Decoupled 3-D Wavelet Decomposition at Spatial Scale Level 2 and Temporal Scale Level 2

3.2 COLOR VIDEO PROCESSING

In color video processing, there are several ways to represent the video signal in color space. The standard format is to represent the data by three primary colors of red, green and blue, but it is not the most efficient way to represent the data for visual purpose. The human visual system has a much smaller dynamic range for spatial variation in color (chrominance) than for brightness (luminance), so an efficient way to encode the data is to use more bits in encoding the luminance data than in encoding the chrominance data. Thus, the luminance channel would be given higher bandwidth than the chrominance channels. Several formats have been defined that use this type of representation of color information, all of them use sets of three values to represent luminance and chrominance components. We will use a popular format, the YUV representation, where the Y-channel contains the luminance information dominated by green, and the (U,V)-channels contain chrominance information which are luminance-subtracted blue and red components, respectively. The latter do not require as many bits to represent as the luminance data. In order to reduce bits assigned to chrominance data, the YUV format in color videos employs different spatial resolution subsamplings. Three such examples, among others, are:

- 4:4:4 – for every 2x2 pixels of luminance, 4 pixels are used to represent each chrominance channel. i.e. no subsampling.
- 4:2:2 – for every 2x2 pixels of luminance, 2 pixels are used to represent each chrominance channel. i.e. subsampling by $\frac{1}{2}$ horizontally.
- 4:1:1 – for every 2x2 pixels of luminance, 1 pixel is used to represent each chrominance channel. i.e. subsampling by $\frac{1}{2}$ both horizontally and vertically.

The transformation from RGB to YUV is:

$$Y = 0.229R + 0.587G + 0.11B;$$

$$U = -0.147R - 0.289G + 0.437B;$$

$$V = 0.615R - 0.515G - 0.1B;$$

Even though luminance is visually more important to the human eyes, data loss in chrominance may also result in the loss of important visual clues as illustrated below in Figure 10. A single video frame without data loss is compared with the same video frame with data loss in either chrominance U or V domains. In this instance, the video frame with data loss in the chrominance U domain exhibits unpleasant yellowish tint, while the video frame with data loss in the chrominance V domain has lost some important information since the dates on the calendar are no longer as distinct as those in the original video frame. So it is important to minimize data loss in both luminance and chrominance domains, even under low bandwidth requirement.

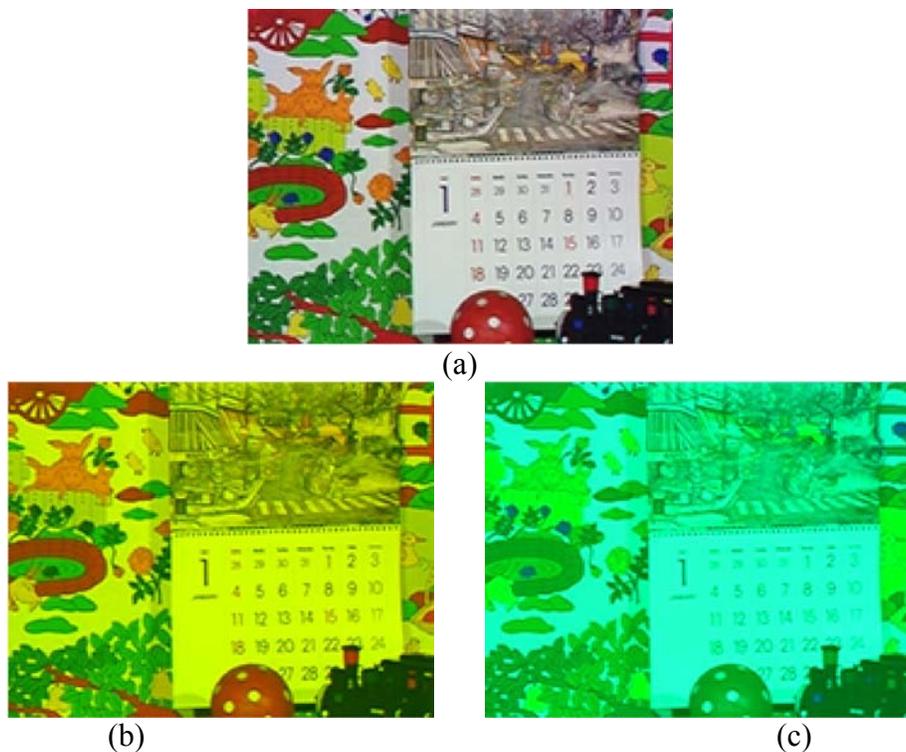


Figure 10 (a) A single frame from Mobile video sequence, (b) the same frame with data loss in chrominance U, and (c) the same frame with data loss in chrominance V

In general, there are two methods to encode wavelet-based color videos. One method is to encode luminance and chrominance wavelet coefficients separately and then insert the encoded data into predetermined locations in the output bitstream independently [61,62]. This method regards the luminance and chrominance data as completely independent, and does not take any advantage of any correlation between them; but it is simple to implement and lends itself easily to parallel processing. The other method is to regard the luminance and chrominance as not completely uncorrelated, and combine their wavelet coefficients into a single representation before encoding [59,64-66]; so this method may take advantage of any correlation between luminance and chrominance, but its complexity is greater than the separate coding method. A particular wavelet used in a wavelet-based video coder should provide good performances in both luminance and chrominance domains. Multiwavelet, having multiple useful features, may provide the flexibility in attaining this requirement.

3.3 3-D MULTIWAVELET DECOMPOSITION IN COLOR VIDEO SEQUENCES

The separable 3-D discrete wavelet transform is formed by the tensor product of three 1-D discrete wavelet transforms, one applied in each dimension separately, in all three color domains. For symmetric wavelet decomposition, the 3-D data block is decomposed to the same scale level along all three dimensions. For decoupled decomposition, the scale level of decomposition along the temporal dimension may be independent from the scale level of decomposition along the spatial dimensions. A one-level 3-D wavelet decomposition of a group of eight frames is illustrated in Figures 11 through Figure 14. The chrominance decomposition coefficients are shown in color to be distinguished from the luminance coefficients. These eight frames are extracted from the video sequence Coastguard which shows a speedboat accelerating toward a coastguard ship as shown in Figure 11. Figures 12 and 13 show the grouping of the

luminance component of the eight frames into one 3-D data block which is then decomposed using the 3-D wavelet transform. Figure 14 shows the 3-D wavelet decomposition for the luminance as well as the chrominance components of the same video segment. (Note that the video frames are flipped along the Y axis for display purpose only.)



Figure 11 A single frame from the Coastguard video sequence

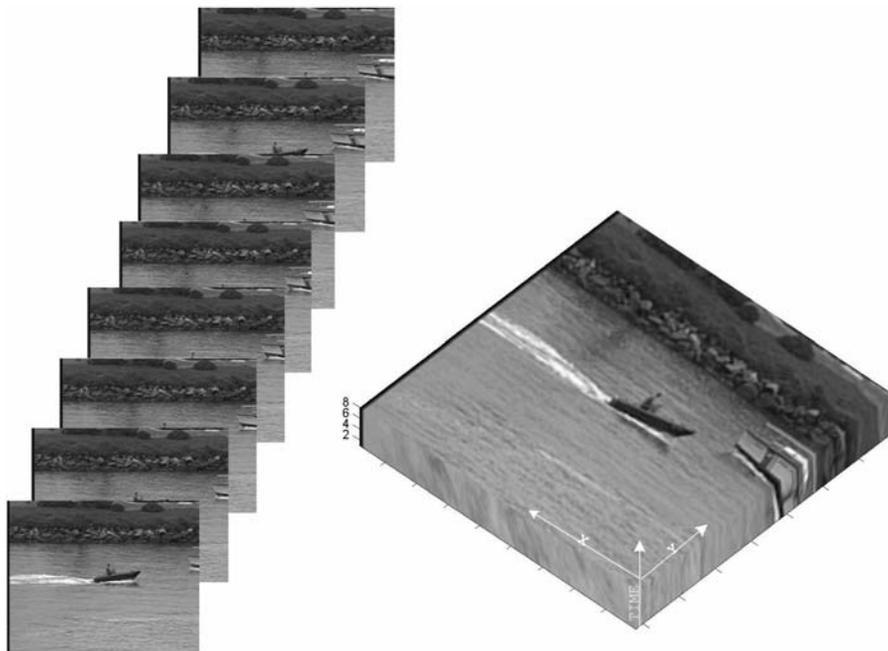


Figure 12 Eight sequential luminance frames from Coastguard grouped into a 3-D block

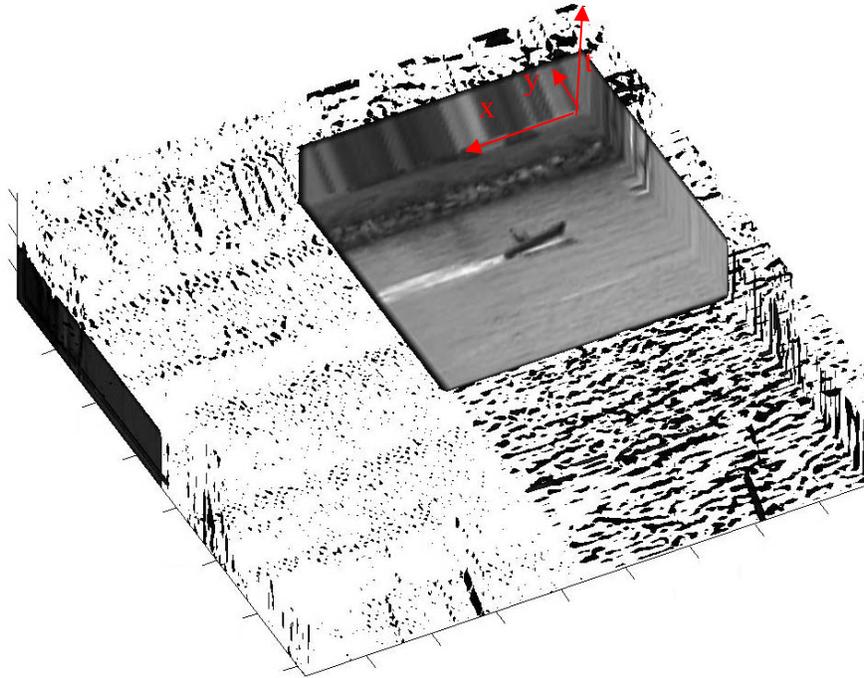


Figure 13 A block of 3-D Wavelet Transform (1-level) of the Luminance Portions from 8 Coastguard Frames

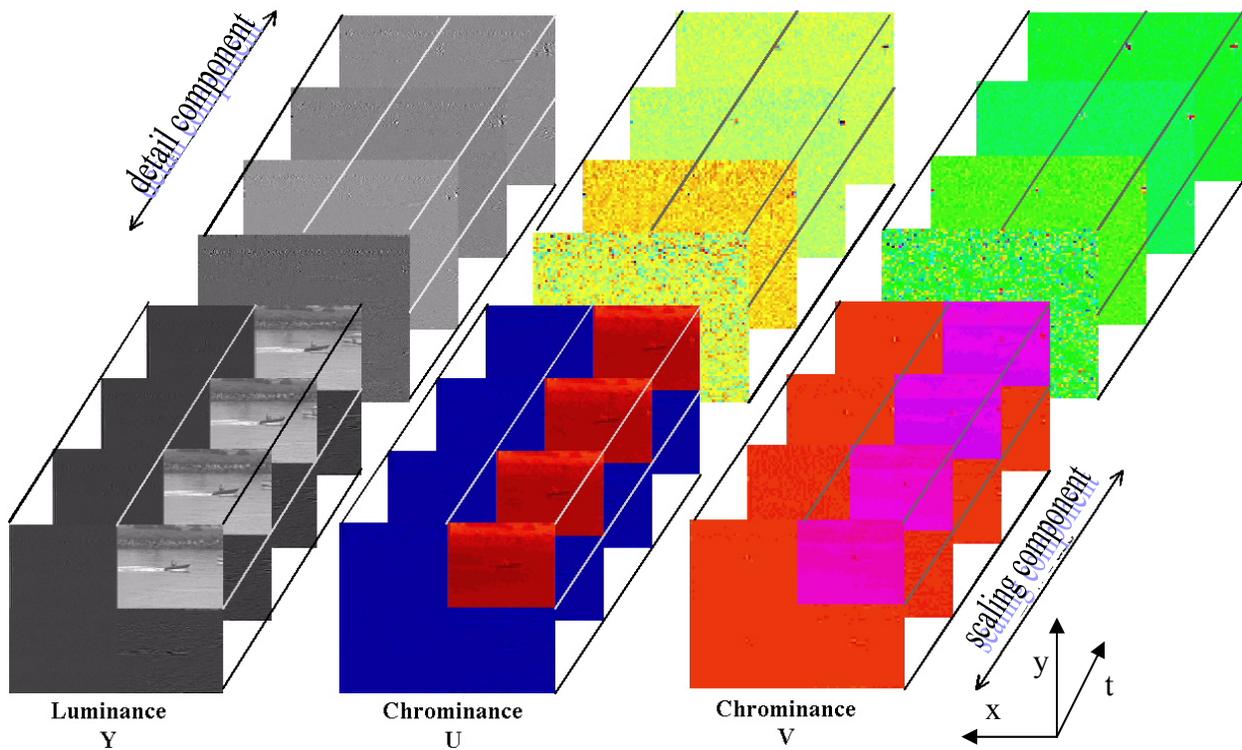


Figure 14 Expanded 3-D Wavelet Blocks of 8 Coastguard Frames in All Three Color Domains

The multiwavelet decomposition of either luminance or chrominance component of color videos is shown in Figure 15 where the three color domains can be considered independent of each other.

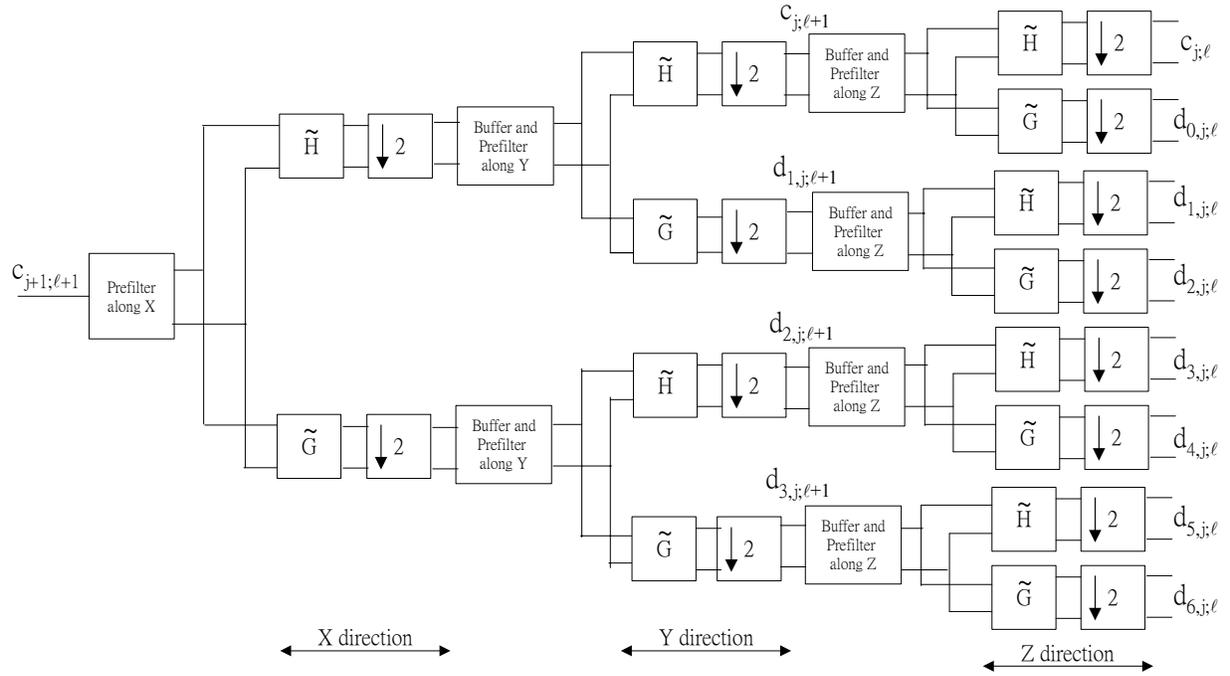


Figure 15 One-level Symmetric Three-Dimensional Multiwavelet ($r=2$) Decomposition along 3 Directions

The prefiltering of the input data, which is required for multiwavelets in general, can be considered outside the decomposition blocks to simplify the mathematical expressions. For the purpose of flexibility, let us first develop the expressions for the case of decoupled decomposition.

Let $\{c_{j; \ell}(n, m; k)\}_{j, \ell \in \mathbb{Z}}$ denote the scaling coefficient matrix at the scale level j of the spatial domain and scale level ℓ of the temporal domain, where n, m and k are spatial and temporal indices, respectively. Similarly, $\{d_{1; j; \ell}(n, m; k)\}$, $\{d_{2; j; \ell}(n, m; k)\}$, and $\{d_{3; j; \ell}(n, m; k)\}$ denote, respectively, the three wavelet coefficient matrices at the spatial scale j and the temporal scale

level ℓ . We will be mainly concerned with the case of $r=2$, $\phi(\mathbf{x}) = \begin{bmatrix} \phi_1(\mathbf{x}) \\ \phi_2(\mathbf{x}) \end{bmatrix}$, $\psi(\mathbf{x}) = \begin{bmatrix} \psi_1 \\ \psi_2 \end{bmatrix}$. In the two-dimensional spatial domain, there is one 2-D scaling function $\phi(\mathbf{x}, \mathbf{y})$ but three 2-D wavelets $\psi_1(\mathbf{x}, \mathbf{y})$, $\psi_2(\mathbf{x}, \mathbf{y})$, and $\psi_3(\mathbf{x}, \mathbf{y})$ for decompositions of the data along horizontal, vertical and diagonal directions,

$$\phi(\mathbf{x}, \mathbf{y}) = \phi(\mathbf{x}) \circ \phi(\mathbf{y})$$

$$\psi_1(\mathbf{x}, \mathbf{y}) = \phi(\mathbf{y}) \circ \psi(\mathbf{x})$$

$$\psi_2(\mathbf{x}, \mathbf{y}) = \psi(\mathbf{y}) \circ \phi(\mathbf{x}), \text{ and}$$

$$\psi_3(\mathbf{x}, \mathbf{y}) = \psi(\mathbf{y}) \circ \psi(\mathbf{x})$$

where \circ denotes the tensor product. The corresponding decomposition filter matrices are:

$$\tilde{H}_{xy}, (n, m) = \tilde{H}_x(n) \tilde{H}_y(m),$$

$$\tilde{G}_{1,xy}, (n, m) = \tilde{G}_x(n) \tilde{H}_y(m),$$

$$\tilde{G}_{2,xy}, (n, m) = \tilde{H}_x(n) \tilde{G}_y(m),$$

$$\tilde{G}_{3,xy}, (n, m) = \tilde{G}_x(n) \tilde{G}_y(m)$$

Then, the 2-D spatial decomposition from scale level $(j+1)$ to scale level j , at a fixed temporal scale $\ell+1$, is given by

$$\bar{c}_{j; \ell+1}(n, m; k) = \sum_{\mu} \tilde{H}_y(\mu - 2m) \sum_{\nu} \tilde{H}_x(\nu - 2n) c_{j+1; \ell+1}(\nu, \mu; k) = \sum_{\mu, \nu} \tilde{H}_{xy}(\nu - 2n, \mu - 2m) c_{j+1; \ell+1}(\nu, \mu; k),$$

$$\bar{d}_{1, j; \ell+1}(n, m; k) = \sum_{\mu} \tilde{H}_y(\mu - 2m) \sum_{\nu} \tilde{G}_x(\nu - 2n) c_{j+1; \ell+1}(\nu, \mu; k) = \sum_{\mu, \nu} \tilde{G}_{1,xy}(\nu - 2n, \mu - 2m) c_{j+1; \ell+1}(\nu, \mu; k),$$

$$\bar{d}_{2, j; \ell+1}(n, m; k) = \sum_{\mu} \tilde{G}_y(\mu - 2m) \sum_{\nu} \tilde{G}_x(\nu - 2n) c_{j+1; \ell+1}(\nu, \mu; k) = \sum_{\mu, \nu} \tilde{G}_{2,xy}(\nu - 2n, \mu - 2m) c_{j+1; \ell+1}(\nu, \mu; k),$$

$$\bar{d}_{3, j; \ell+1}(n, m; k) = \sum_{\mu} \tilde{G}_y(\mu - 2m) \sum_{\nu} \tilde{G}_x(\nu - 2n) c_{j+1; \ell+1}(\nu, \mu; k) = \sum_{\mu, \nu} \tilde{G}_{3,xy}(\nu - 2n, \mu - 2m) c_{j+1; \ell+1}(\nu, \mu; k)$$

$\forall k \in \mathbf{Z}$

The decomposition along the temporal dimension, at the fixed spatial scale j is given by

$$\vec{c}_{j,\ell}(n,m;k) = \sum_{\rho} \tilde{H}_{\ell}(\rho-2k) c_{j,\ell+1}(n,m;\rho), \quad \forall k \in \mathbf{Z}$$

and

$$\vec{d}_{j,\ell}(n,m;k) = \sum_{\rho} \tilde{G}_{\ell}(\rho-2k) c_{j,\ell+1}(n,m;\rho), \quad \forall k \in \mathbf{Z}$$

For the symmetric 3-D decomposition, the same decomposition at level j is used in spatial and temporal domains,

$$\vec{c}_j(n,m;k) = \sum_{\rho} \tilde{H}_z(\rho-2k) \sum_{\mu,\nu} \tilde{H}_{xy}(v-2n, \mu-2m) c_{j+1}(v, \mu, \rho) = \sum_{\mu,\nu,\rho} \tilde{H}_{xyz}(v-2n, \mu-2m, \rho-2k) c_{j+1}(v, \mu, \rho),$$

$$\vec{d}_{0,j}(n,m;k) = \sum_{\rho} \tilde{G}_z(\rho-2k) \sum_{\mu,\nu} \tilde{H}_{xy}(v-2n, \mu-2m) c_{j+1}(v, \mu, \rho) = \sum_{\mu,\nu,\rho} \tilde{G}_{0,xyz}(v-2n, \mu-2m, \rho-2k) c_{j+1}(v, \mu, \rho),$$

$$\vec{d}_{1,j}(n,m;k) = \sum_{\rho} \tilde{H}_z(\rho-2k) \sum_{\mu,\nu} \tilde{G}_{1,xy}(v-2n, \mu-2m) c_{j+1}(v, \mu, \rho) = \sum_{\mu,\nu,\rho} \tilde{G}_{1,xyz}(v-2n, \mu-2m, \rho-2k) c_{j+1}(v, \mu, \rho),$$

$$\vec{d}_{2,j}(n,m;k) = \sum_{\rho} \tilde{G}_z(\rho-2k) \sum_{\mu,\nu} \tilde{G}_{1,xy}(v-2n, \mu-2m) c_{j+1}(v, \mu, \rho) = \sum_{\mu,\nu,\rho} \tilde{G}_{2,xyz}(v-2n, \mu-2m, \rho-2k) c_{j+1}(v, \mu, \rho),$$

$$\vec{d}_{3,j}(n,m;k) = \sum_{\rho} \tilde{H}_z(\rho-2k) \sum_{\mu,\nu} \tilde{G}_{2,xy}(v-2n, \mu-2m) c_{j+1}(v, \mu, \rho) = \sum_{\mu,\nu,\rho} \tilde{G}_{3,xyz}(v-2n, \mu-2m, \rho-2k) c_{j+1}(v, \mu, \rho),$$

$$\vec{d}_{4,j}(n,m;k) = \sum_{\rho} \tilde{G}_z(\rho-2k) \sum_{\mu,\nu} \tilde{G}_{2,xy}(v-2n, \mu-2m) c_{j+1}(v, \mu, \rho) = \sum_{\mu,\nu,\rho} \tilde{G}_{4,xyz}(v-2n, \mu-2m, \rho-2k) c_{j+1}(v, \mu, \rho),$$

$$\vec{d}_{5,j}(n,m;k) = \sum_{\rho} \tilde{H}_z(\rho-2k) \sum_{\mu,\nu} \tilde{G}_{3,xy}(v-2n, \mu-2m) c_{j+1}(v, \mu, \rho) = \sum_{\mu,\nu,\rho} \tilde{G}_{5,xyz}(v-2n, \mu-2m, \rho-2k) c_{j+1}(v, \mu, \rho),$$

$$\vec{d}_{6,j}(n,m;k) = \sum_{\rho} \tilde{G}_z(\rho-2k) \sum_{\mu,\nu} \tilde{G}_{3,xy}(v-2n, \mu-2m) c_{j+1}(v, \mu, \rho) = \sum_{\mu,\nu,\rho} \tilde{G}_{6,xyz}(v-2n, \mu-2m, \rho-2k) c_{j+1}(v, \mu, \rho)$$

These 3-D multiwavelet decompositions are initiated by applying them first to the prefiltered luminance and prefiltered chrominance data separately. The decompositions can then

be carried out from one scale level to the next. The SPIHT algorithm originally developed by Said and Pearlman[60] for encoding 2-D images can be extended to encode such 3-D scaling coefficients and wavelet coefficients.

3.4 3-D SPIHT IN SPATIO-TEMPORAL 3-D WAVELET DOMAIN

Said and Pearlman developed a method of encoding data with built-in hierarchy which they called Set-Partitioning-In-Hierarchical-Tree (SPIHT). The key feature of SPIHT is that it automatically generates efficient embedded bitstreams without the additional use of arithmetic coders. In the embedded bitstreams, the information is automatically organized in bits such that the most important bits are sent first so that the bitstream can be cutoff at any point and the best possible data can still be reconstructed from the lossy data.

The basic principle of the SPIHT is briefly reviewed below: Given a set of data with an integrated hierarchy, define a spatial orientation tree structure with parent-children relationship within the data using its hierarchy. The simplest notion of an hierarchical relationship between a parent and its children is that if the value of a parent is significant according to some threshold, then at least one of its children should also be significant relative to the same threshold. The parent-children relation for 2-D data with two levels of hierarchy is shown in Figure 16, and as illustrated, each parent node either have three children if it in the coarsest subband, four children if it is in one of the intermediate subbands, or no child at all.

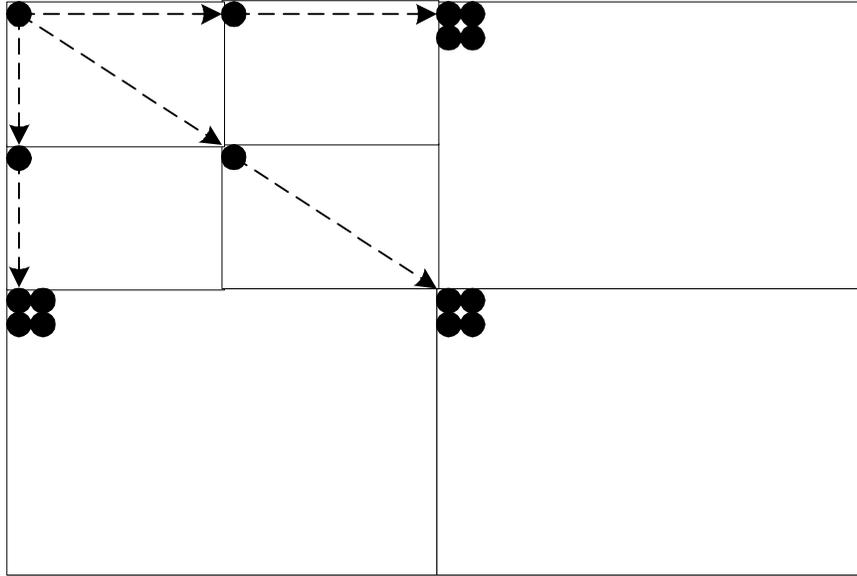


Figure 16 2-D SPIHT Tree Structure with Dashed Lines indicating Parent-Children Relationships

The wavelet decomposition coefficients are scanned in multiple passes and selectively encoded, and under such a tree structure the maximum number of passes required is determined by the number of bits needed to represent the largest coefficient. Let us assume that this number is n , so the threshold used to determine significance during the first pass is set at 2^{n-1} , and this threshold is decremented in each successive pass by a power of 2. In order to track the significance of each coefficient, SPIHT partitions the coefficients and the linking trees into three categories: significant pixels/coefficients, insignificant sets and insignificant pixels/coefficients; anytime a coefficient is moved into the significant category from the others, a bit representing its sign is sent. In each pass two operations take place: a sorting phase and a refinement phase; the bit value of each significant pixel/coefficient at the location just above the current threshold is sent in the sorting phase, and then each set and pixel/coefficient in the insignificant category is examined against the threshold to see if any data point can be moved into the category of significant pixels/coefficients in the refinement phase. This process repeats itself until all the

available bits are sent. In this algorithm, the main saving is that only a single bit is required to represent the case that all the descendents of a node are insignificant at a given threshold value. The SPIHT encoder is a very efficient encoder if the relative significance of parent-children relationships holds true for every node in the tree.

The basic algorithm is given as follows below [60]:

0. Initialize three linked lists:

LIP: List of Insignificant Pixels (or coefficient)

LIS: List of Insignificant Sets

(The insignificant set associated with an insignificant pixel or coefficient in LIP is a sub-tree originating from the significant pixel. A tree may be of type D which include all children and all generations below or it could be of type L which includes all grandchildren and below but not the immediate children)

LSP: List of Significant Pixels (pixels above the current testing threshold)

1. Send the number of bits required to code the absolute value of the largest coefficient as 'n'
2. Extract the pixels (coefficients) from the root node (LL) and insert them in LIP. Insert the trees associated with the root node pixels into LIS as type D. Set LSP to be an empty set.
3. Set the threshold as $2^{n-1}-1$ so that all the coefficients whose absolute values require n binary bits to represent will be above the threshold and be significant, then start the sorting pass.
4. Sorting pass:
 - I. Check all the LIP elements against the threshold, and move all the significant elements from LIP to LSP after sending a '1' and its sign bit. A '0' is sent for each insignificant element in the LIP.
 - II. Check all the LIS elements for significance.

- For a 'D' type tree, send a '1' if one of the children or grandchildren and below is significant, then check each child against the threshold. If a child is significant, output a '1' plus its sign and move it to LSP set, but if a child is not significant, output a '0' and move it to the LIP. If grandchildren exist, move this tree to the end of LIS as type L, otherwise remove it from LIS. A '0' is sent if the entire D tree is insignificant.
- For an 'L' type tree, send a '1' if one of the grandchildren and below is significant, remove the originating node from LIS and add its children to the end of LIS as type 'D', otherwise send a '0'.

5. Refinement pass:

- Check each of the original elements in the LSP, which are elements that existed in LSP before the sorting pass, against the threshold and output a '1' if significant, and a '0' if otherwise.

6. Reduce the testing threshold, 'n', by 1 bit, and if $n > 0$, then repeat from step 3.

Several versions of 3-D SPIHT, with different parent-children such as those shown in Figures 17 and 18, have been explored [57-66] with varying degrees of success depending on the input sequence. Once a specific parent-children relationship that can be extended to 3-D SPIHT structure is defined, the above algorithm for the 2-D SPIHT can be extended to 3-D SPIHT by initializing the LIP and LIS using the coefficients in the LLL subband as defined in Figures 8 and 9, respectively. The tree structure for the symmetric decomposition, shown in Figure 17, was used in this study.

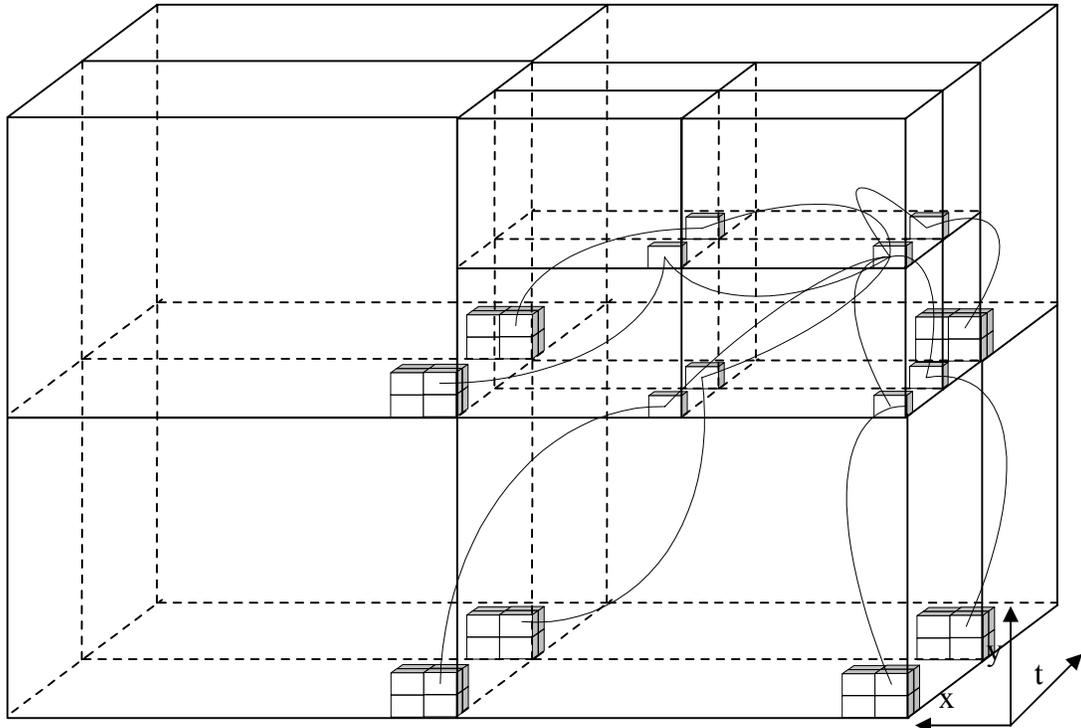


Figure 17 Symmetric 3-D SPIHT Tree Structure with Curved Lines indicating Parent-Children Relationships

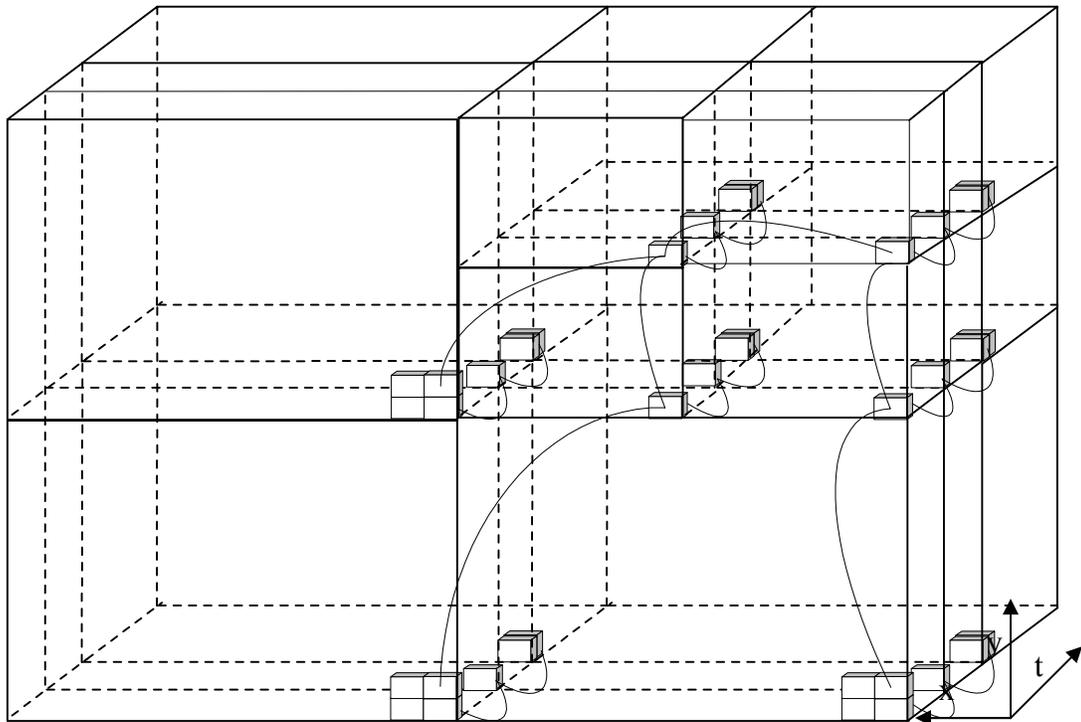


Figure 18 Decoupled 3-D SPIHT Tree Structure with curved lines indicating Parent-Children Relationships

3.5 EXPERIMENTS AND RESULTS

This section describes the experiments set up to test the three-dimensional multiwavelet transform based against the three-dimensional single wavelet for video compression. Because the framework of the experiments is for video compression, the prefiltering scheme chosen for the experiments is the “critically-sampled approximation” scheme so as to economize preprocessing. The type of 3-D wavelet decomposition chosen was the symmetric decomposition, and the level of (multi)wavelet decomposition was mostly limited to be two. The latter was due to the use of relatively short video segments in order to reduce the inclusion of frames containing zooming, panning and rotation.

To make sure that the performance evaluation is valid across a wide spectrum of different video types, 64 video frames were extracted from each of nine different QCIF-sized (144x176 pixels) color video sequences in 4:1:1 YUV format: “Claire”, “Miss America”, “Foreman”, “Silent”, “Coastguard”, “Mobile”, “Salesman”, “Stefan”, and “Table Tennis”, one frame each is shown in Figure 19. These video sequences, with different contents and motions, represent a wide spectrum of possible video sources such as videoconference, surveillance, advertisement, educational programming, sports event, etc.



Figure 19 Video frames from (a) Claire, (b) Miss America, (c) Foreman, (d) Silent, (e) Coastguard, (f) Mobile, (g) Salesman, (h) Stefan, and (i) Table Tennis sequences.

Two multiwavelets and a single wavelet with similar properties were chosen in the experiment. They are the GHM multiwavelet, Chui-Lian multiwavelet and the Daubechies 4-tapped single wavelet. For prefiltering the data for multiwavelet decomposition, the approximation prefiltering coefficients for GHM and Chui-Lian were computed using the algorithm described by Strela [9]. The approximation prefilter $Q(n)$ and post-filter $P(n)$ coefficients for GHM multiwavelet are:

$$Q_0 = \begin{bmatrix} \frac{3}{16} & \frac{10}{16} \\ 0 & 0 \end{bmatrix}, Q_1 = \begin{bmatrix} \frac{3}{16} & 0 \\ \frac{1}{\sqrt{2}} & 0 \end{bmatrix}$$

$$P_0 = \begin{bmatrix} 0 & \frac{1}{\sqrt{2}} \\ 0 & \frac{-3}{10\sqrt{2}} \end{bmatrix}, P_1 = \begin{bmatrix} 0 & 0 \\ \frac{4}{5} & \frac{-3}{10\sqrt{2}} \end{bmatrix}$$

and the approximation prefilter coefficients and post-filter coefficients for Chui-Lian multiwavelet are:

$$Q = \begin{bmatrix} \frac{1}{4\sqrt{2}} & \frac{1}{4\sqrt{2}} \\ \frac{1}{(1+\sqrt{7})\sqrt{2}} & \frac{-1}{(1+\sqrt{7})\sqrt{2}} \end{bmatrix}, P = \begin{bmatrix} \frac{2}{\sqrt{2}} & \frac{1+\sqrt{7}}{2\sqrt{2}} \\ \frac{2}{\sqrt{2}} & \frac{-1-\sqrt{7}}{2\sqrt{2}} \end{bmatrix}$$

The performance comparison was done using by using a symmetric three-dimensional SPIHT as the encoder. Four low bitrates were tested for the initial experiments: 0.1 bit-per-pixel, 0.25 bit-per-pixel, 0.33 bit-per-pixel and 0.5 bit-per-pixel, in order to examine how high levels of quantization affect the performance of multiwavelet-based 3-D compression.

Given the color video sequences listed above, each frame within a sequence is decomposed into its Y,U, and V color components initially. Then sixteen frames of data from each component are grouped together into a 3-D data block and then decomposed using symmetric multiwavelet decomposition to level 2. The resulting coefficients are converted to an embedded bitstream using the Symmetric 3-D SPIHT at the bit-rates specified above. Each resultant bitstream is then decoded and reconstructed back into the spatial domain in order to obtain lossy versions of the original video frames in YUV format. For objective measurements of the qualities of these frames against the original video frames, the PSNR measurement is used for each color component of each video frame. Figure 20 illustrates the entire process.

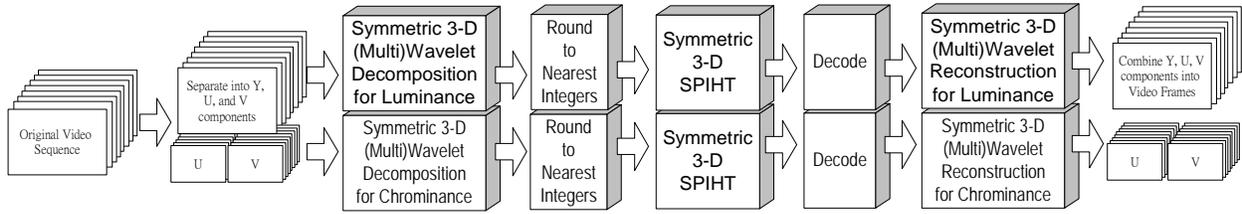


Figure 20 Experimental Setup of 3D (Multi)Wavelet-based Color Video Compression

Due to the nature of the rounding operator as a quantizer, the possibility exists that the SPIHT-encoded result may be appropriate in the sense that rounding operator in combination with the SPIHT quantizer will result in bitstreams shorter than the target bitrate; i.e. the length of the bitstream required to code perfectly the input to the 3-D SPIHT encoder is less than the target bitrate multiplied by the total number of pixels. In the experiment, any result that matched with this condition was considered invalid and were not included in the results for discussion.

From the results of this experiment, some advantages of applying multiwavelets to three-dimensional color video compression have been observed. At the very low bit-rate of 0.1 bit-per-pixel, the PSNR measurements of almost all of the videos encoded using 3-D multiwavelets, (the Chui-Lian multiwavelet), visibly exceeded the PSNR measurements of the same videos encoded using 3-D single wavelet (DB4) by a considerable amount in all three color domains. This is illustrated in Figure 21 for the Coastguard video sequence shown below. (See Appendix for detail results.)

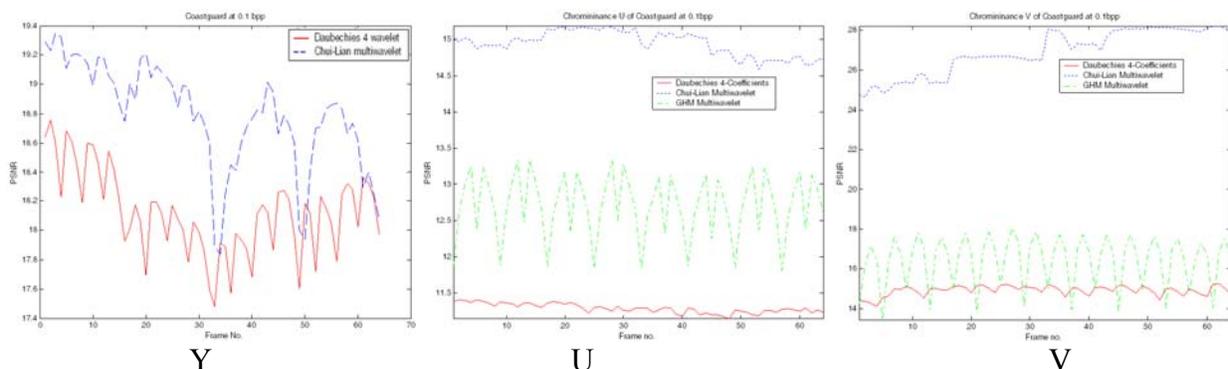
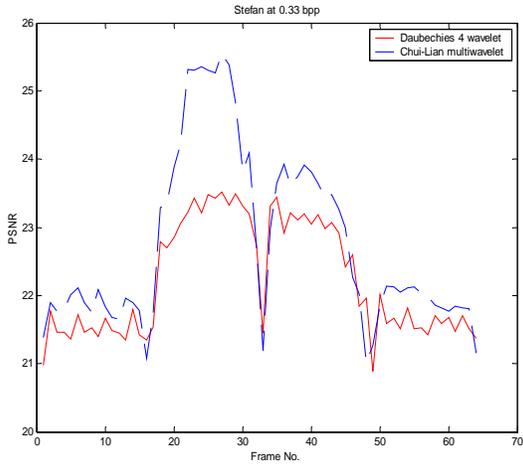
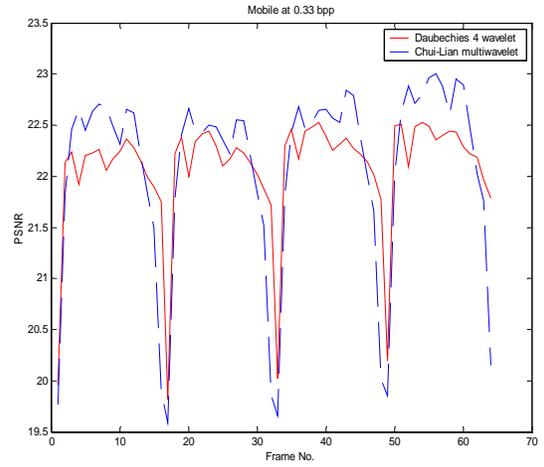


Figure 21 Spatio-Temporal 3-D Video Compression: PSNR Measure of Coastguard Video Sequence at at 0.1 bpp

Since the high compression ratio of 0.1 bit-per-pixel in SPIHT encoder implies that only the most significant scaling coefficients were used in the reconstruction, the observation of better performance of multiwavelets gives weight to the proposition that multi-scaling functions can provide a better approximation of the video than a single scaling function. In the very low-bit-rate of 0.1 bit-per-pixel, multiwavelets offer obvious advantages in vast majority of the video sequences tested regardless of lighting conditions, motions of the objects in the video or motions of the cameras. It is observed that these factors do affect the results when bit-rates are more moderate; the results of 0.33 bit-per-pixel compression of the luminance portions of Stefan and Mobile video, shown below in Figure 22, do not show as obvious advantage for using multiwavelet over single wavelet. These two video sequences contain both panning and zooming motions which implies that video frames within each group are not well correlated with each other. But when the zooming and panning motions paused in the second group of frames of the Stefan video sequence, the advantages of using multiwavelets were revealed again. The results of 3-D compression of the Salesman video at 0.33 bit-per-pixel is shown in Figure 23, offers a good illustration. This video sequence contains only simple motions within the chosen groups of frames. Based on these observations, it is concluded that, under moderate bit-rates, the multi-scaling functions of 3-D multiwavelet systems can better approximate video frames that are more correlated with each other than the scaling function in single wavelet system.

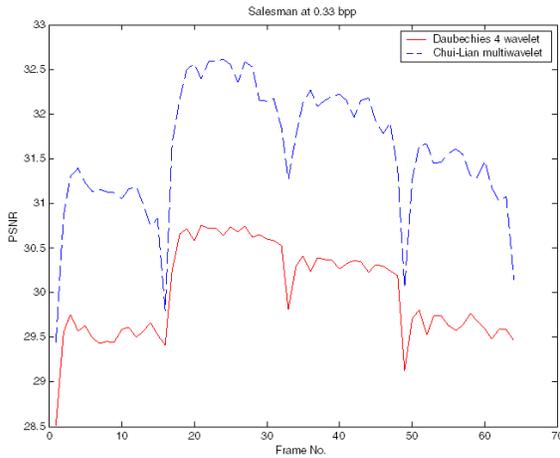


(a)

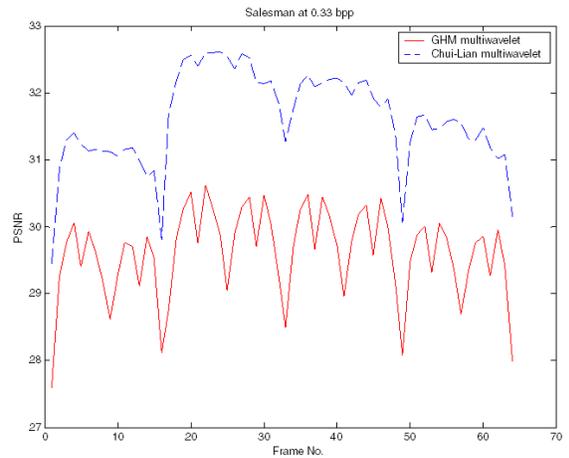


(b)

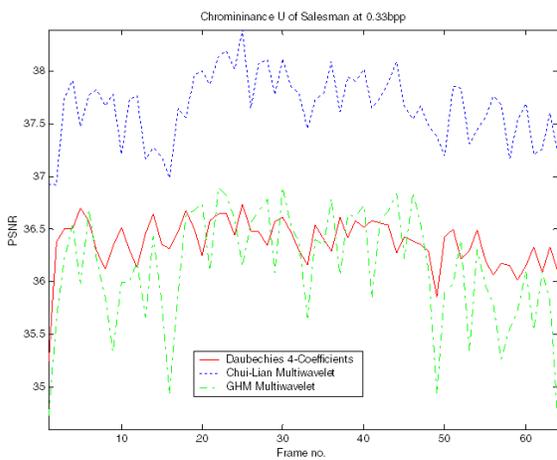
Figure 22 Spatio-Temporal 3-D Compression using CL multiwavelet and DB4 PSNR at 0.33 bpp for luminance portions of (a) Stefan, and (b) Mobile



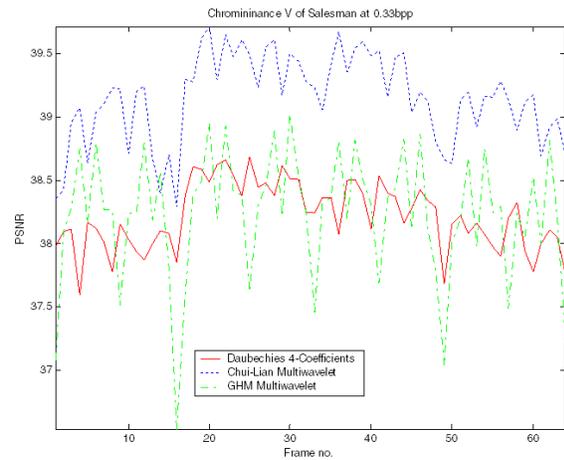
Luminance: CL vs. DB4



Luminance: CL vs. GHM



Chrominance U: CL, GHM, and DB4



Chrominance V: CL, GHM, and DB4

Figure 23 Spatio-Temporal 3-D Compression using CL, GHM Multiwavelets and DB-4, PSNR at 0.33 bpp for Salesman

From these observations derived from the results of one experiment under 2-level symmetric 3-D multiwavelet decompositions, we reached the conclusion that the Chui-Lian multiwavelet is better suited to three-dimensional wavelet coding than the comparable single wavelets, especially under the demands of high compression. Although the advantages of multiwavelets are not as obvious in all cases when the compression ratios are more moderate, most of the results do show that the Chui-Lian multiwavelet exhibits obvious advantages.

Another interesting observation should be noted here concerning color videos: that the color domains of the same video sequence may responded in terms of compression. Specifically, in the Miss America video sequence coded at 0.25 bit-per-pixel as shown in Figure 24, the PSNR measurements for the first three groups of frames displays obvious advantage for the Chui-Lian multiwavelet in all domains by varying amount. In the fourth group of pictures, however, when the subject in the video suddenly shifted her position, the PSNR measurements in the case of the Chui-Lian multiwavelet dropped sharply such that it even had a slight loss when compared to the measurements obtained in the single wavelet case in the luminance domain. But no sharp drop in PSNR was observed in both chrominance domains. This suggests that if the luminance and chrominance coefficients are encoded together, then the qualitative loss in one domain may be compensated by the improvement gained in other domains such that the available bits may be allocated in such a way so as to improve the overall quality.

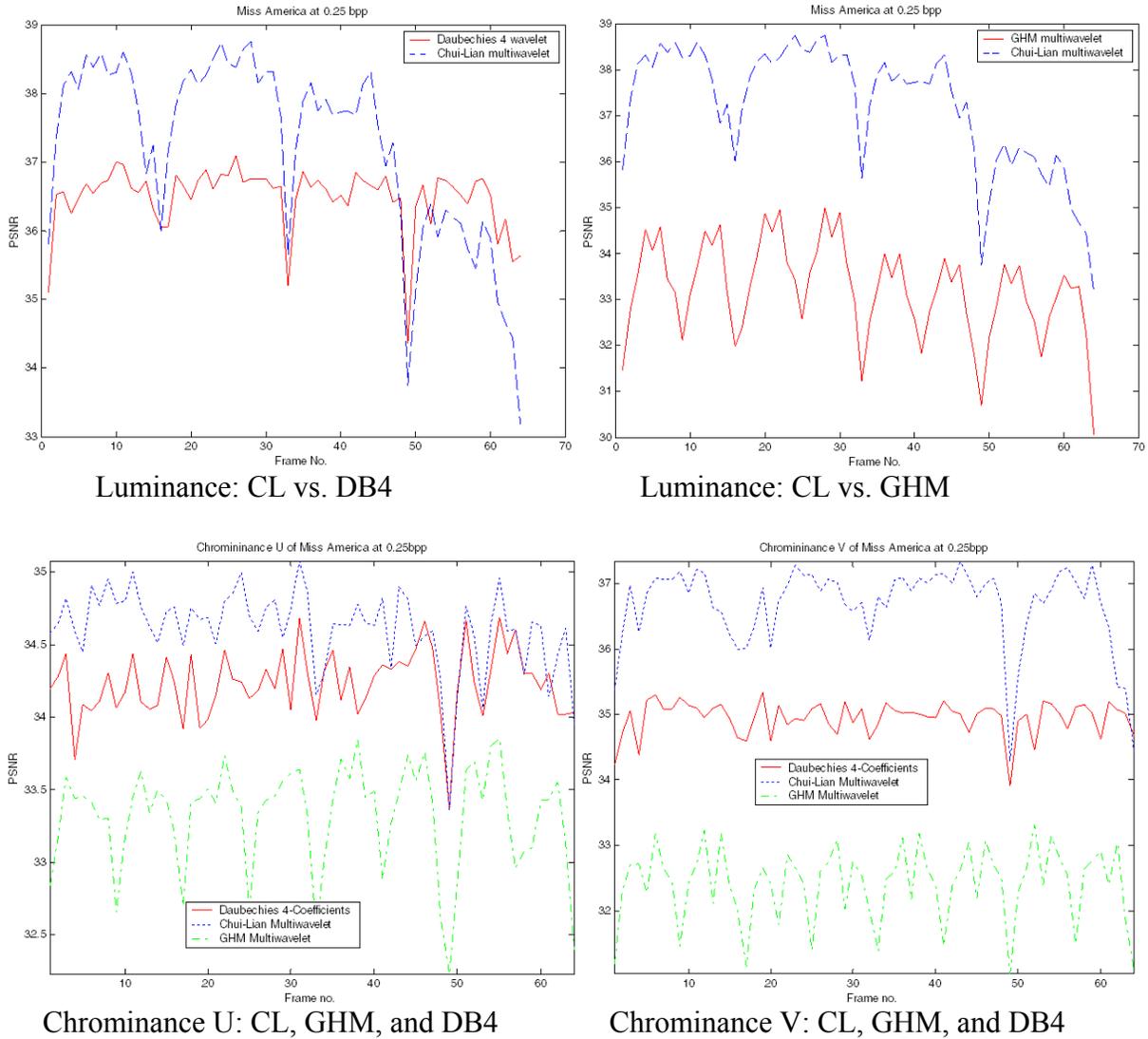


Figure 24 Spatio-Temporal 3-D Compression using Multiwavelets (CL and GHM) and DB4, PSNR at 0.25 bpp for Miss America

3.6 SUMMARY

We conclude that the Spatio-Temporal 3-D multiwavelet video coder performed better than the 3-D single wavelet video coder at low bit-rates between 0.1bpp (compression ratio 1/80) and 0.25bpp (compression ratio 1/32). Under the 2-level symmetric decomposition, the multiwavelet CL, on the average, outperformed the single wavelet DB4 by more than 1dB in the luminance

domain, by 4.5 dB in the chrominance U domain and by 5dB in the chrominance V domain at 0.1bpp. At the bit-rate of 0.25bpp, multiwavelet CL showed a gain in the range of 0 to 2dB in the luminance domain, -0.5dB to 2dB in the chrominance U domain and 0.5dB to 4dB in the chrominance V domain. The gain decreases, however, as the compression ratio increases.

4.0 3-D MULTIWAVELET-BASED COARSE TRAJECTORY RESIDUALS VIDEO CODING

In standard video coding it is generally assumed that the intensity of an object point p , $I(p)$, does not change significantly along the trajectory of its movements within a given short time interval. Given this assumption, we calculate the trajectory of an object's movement from the current frame in which it is occupying to the next frame in time by matching the intensity value of all possible locations in the next frame against the object's current intensity value as illustrated in Figure 25. If there exists no exact match but the closest match in intensity is picked, the difference between these two intensity values is called the motion compensation residual, $r_{j+1}(p) = |I_j(p) - I_{j+1}(p)|$. Encoding motion vectors and small residuals leads to coding efficiency as desired.

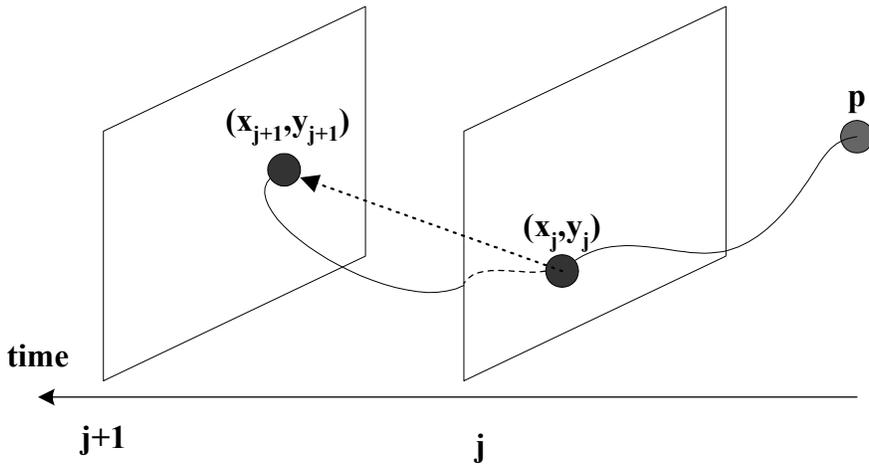


Figure 25 Object P Travels in Successive Video Frames

In the previous chapter, we employ 3D multiwavelets to encode intensity values of pixels of a group of video frames. In this chapter, we consider the use of 3D multiwavelet to the coding

of residuals over time. Temporal redundancy would be removed twice, once through the motion estimation/motion compensation, and second time, through the wavelet decomposition of the spatio-temporal residual function.

4.1 COARSE MOTION RESIDUALS VIDEO CODING

In the-dimensional video coding, only a single frame called a key frame within a video segment is coded using its original intensity values, and all subsequent frames in the segment are coded using the motion vectors and prediction errors (residuals) with reference to the key frame. The three-dimensional multiwavelet coding can also be applied to the residual $r(p,t)$ as a function of (x,y,t) . But in lossy coding, video frames coded using this approach will lose information over time. For example, the residuals for the standard lossless video compression model of any point p can be defined as:

$$r_i(p) = I_i(p) - I_{i-1}(p), \text{ for } i-1 > 0$$

at time i . However, in the case of lossy coding, the process of coefficients quantization will degrade the intensity value of original video frame, so errors will be introduced to the motion compensation residuals:

$$r'_i(p) = I_i(p) - I_{i-1}(p) + e_{i,i-1}(p), \text{ for } i-1 > 0$$

and the new residual value $r'_i(p)$ is not just $r_i(p)$, but $r_i(p)$ plus errors introduced by the process of quantization. Without corrections by additional information, these quantization errors will propagate and accumulate over time. Although the model for the new residuals will grow in complexity over a large number of frames but if we limit the total number of frames, n , to be a small value, then we can approximately model the residuals from lossy compression as:

$$r'_i(p) = I_i(p) - I_{i-1}(p) + e_{i,i-1}(p) + e_{i-1,i-2}(p) + \dots + e_{i-n+1,i-n}(p) \approx r_i(p) + e_{i,i-n}(p)$$

Let us call the smallest number of frame which the residual model above holds true for any given video sequence X as n_X .

From the above residual model, we know that errors will propagate from frame to frame. One method most two-dimensional video coders used to reduce the propagation of errors is the introduction of a feedback loop of each frame after quantization, so that the motion trajectories are re-evaluated from the quantized video frame rather than from original video frames which is illustrated in Figure 26. However, because the residuals are evaluated from quantized video frames, the model above for the residuals for lossy compression holds true for this model.

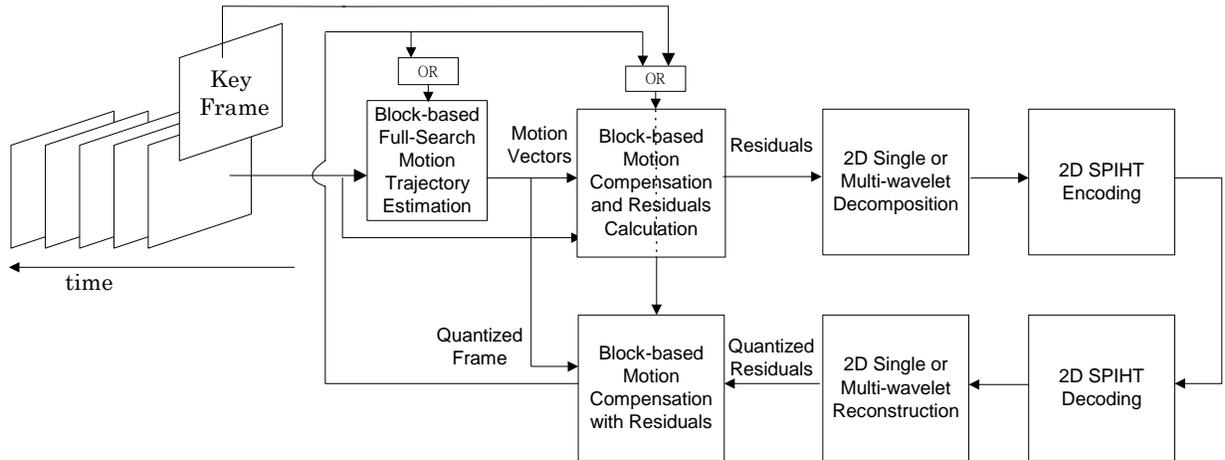


Figure 26 The Block Diagram of Two-dimensional Wavelet-based Lossy Video Compression

Because 3D video coding can not provide a frame-by-frame feedback loop within a given group of frames that are coded together, there exists no common point for a direct comparison of 3D models with the above 2D model. In order to facilitate comparison an intermediary model is introduced that contains aspects of both two-dimensional and three-dimensional video coding but only codes only n_X frames of any given video sequence X . In this intermediary model, like the two-dimensional video model, nearest-neighbor motion trajectory estimation is performed on each frame successive to the key frame but without the use of a frame-by-frame feedback loop.

The residuals are then accumulated in order to form a 3D block for three-dimensional multiwavelet decomposition. The encoding process of this intermediary model is as shown in the figure below:

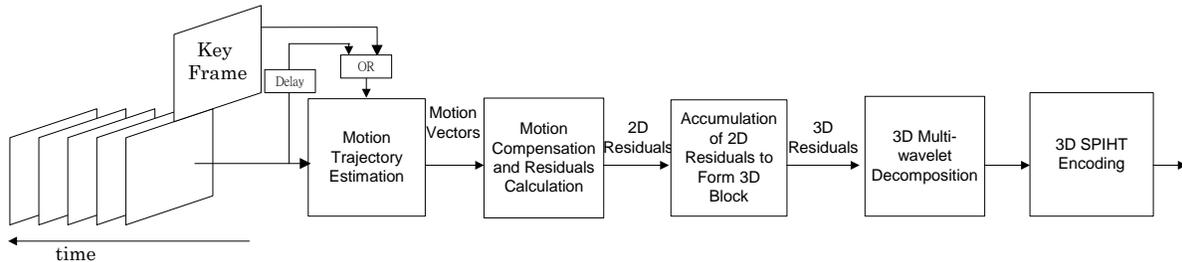


Figure 27 Intermediary Model for Lossy Coding of Multiwavelet-based 3-D Video Compression

The decoding process for the intermediary is the reverse of the encoding process: where the three-dimensional coefficient block is reconstructed to the original time domain then added back to the motion reconstructed frames. However, since the original frames are no longer available, it is assumed that each reconstructed frame approximates its corresponding original frame and uses it to reconstruct the successive frame. This process is illustrated in Figure 28 below:

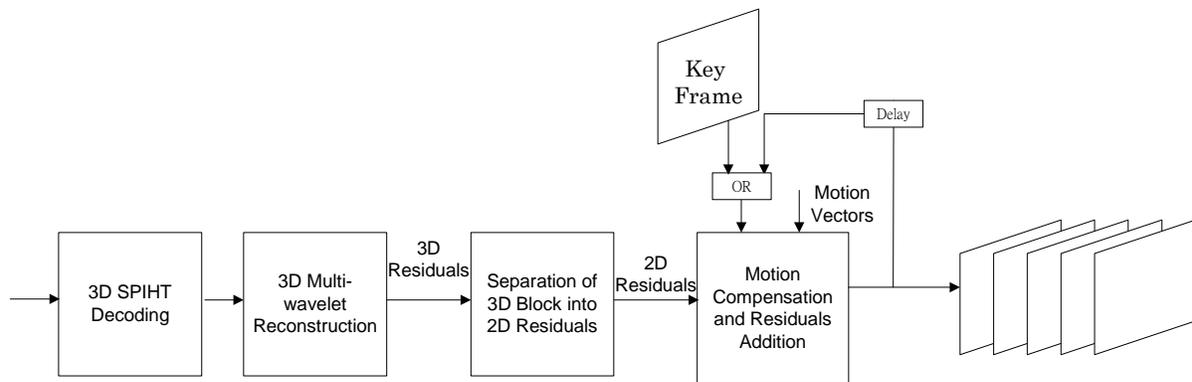


Figure 28 Intermediary Model for Lossy Decoding 3-D Multiwavelet-based Video

The intermediary model is not suitable for video compression; however, it can be useful as a tool of comparison between two-dimensional and three-dimensional models of lossy compression.

For the purpose of simplifying the comparison, we assume that 2D multiwavelet compression is equally effective as 3D multiwavelet compression for these n_X frames so that we do not need to factor in their relative performance into consideration now.

By comparing the residuals that will be coded by both models for n_X frames, it can be noted that the residuals for the intermediary three-dimensional model are smaller since it does not need to carry the burden of coding the extra errors, $e_{i,i-n}(p)$, that the two-dimensional model needs to handle; so the performance of the intermediary model should be superior to two-dimensional video coding for n_X frames. It is our objective in this chapter to develop a third model that would have performance approaching the intermediary model for frames nearest to the key frame yet maintains an average performance above the 2D model for a large number of frames.

There are several proposed models in the literature which combines two-dimensional motion estimation and three-dimensional single wavelet video coding [71, 72, 73 75], but since these methods code the frames using their original intensity values rather than the residuals, so it is difficult to compare them directly with the two-dimensional motion residual coding on an equal platform except by the end results of compression. An example of such a method is the block-based motion threading 3D wavelet transform proposed by Xu, et.al [71], where the non-overlapping motion trajectories are evaluated for each pixel which are used to align them in all the frames temporally in order to increase the efficiency of 3D wavelet transform by applying the transform in the third dimension along the “motion threads” rather than along the dimension of time, as shown in the figure below.

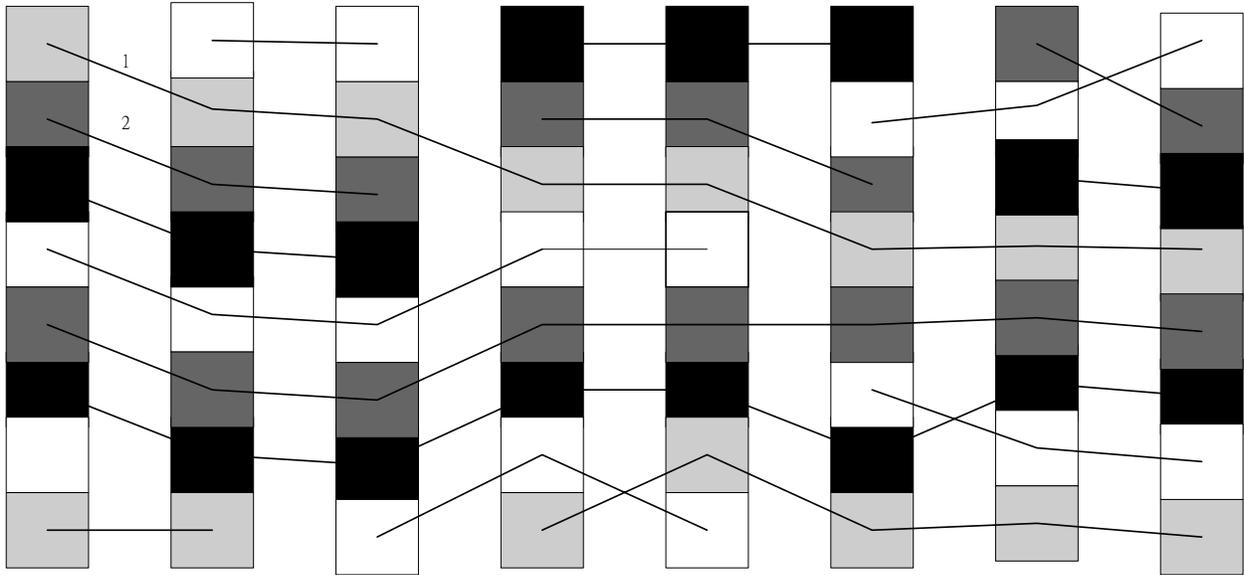


Figure 29 Threads Describing Motion Trajectories of pixels within a Block in a Video Sequence

One of the drawbacks of computing motion threads is that the third dimension of three-dimensional transform is uneven; note in the Figure 29 above, thread 1 threaded through the entire 8 frames, while thread 2 threaded only 3 frames. This drawback implies both that the motion threads must be tracked at pixel-level due to the non-overlapping motion requirement which lacks the flexibility of many-to-one mapping, and that the level of wavelet decomposition in the third dimension for each thread will be different for each thread because each string (pixels sharing the same thread) must be transformed separately from other strings before encoding. The pixel-level motion tracking of motion threads requires greater computational complexity than block-based motion tracking, and different levels of decomposition along the third dimension will introduce visible boundary effects at low bitrates. Luo, et.al [72] sought to reduce some effect by using a complex multilayered lifting-scheme based motion threading, but problems still persist at low bitrates.

Since our target operation is at low bitrates, we propose the common reference model, which is the idea that instead of each frame referenced to its nearest frame and built up the temporal dependence described above; the frames within the same block all refer to the same key frame, chosen from amongst them, in the calculation of motion trajectory residuals. In comparison to the intermediary model, the motion trajectory of each point calculated from the common reference model can be considered as equivalent to the motion trajectory calculated from successive frames. In Figure 30, the dashed lines indicate trajectories evaluated by the intermediary model and the solid lines indicate coarse motion trajectory to the key frame. It can be seen that these two sets of motion information is equivalent since one set of vectors can be obtained from the other set.

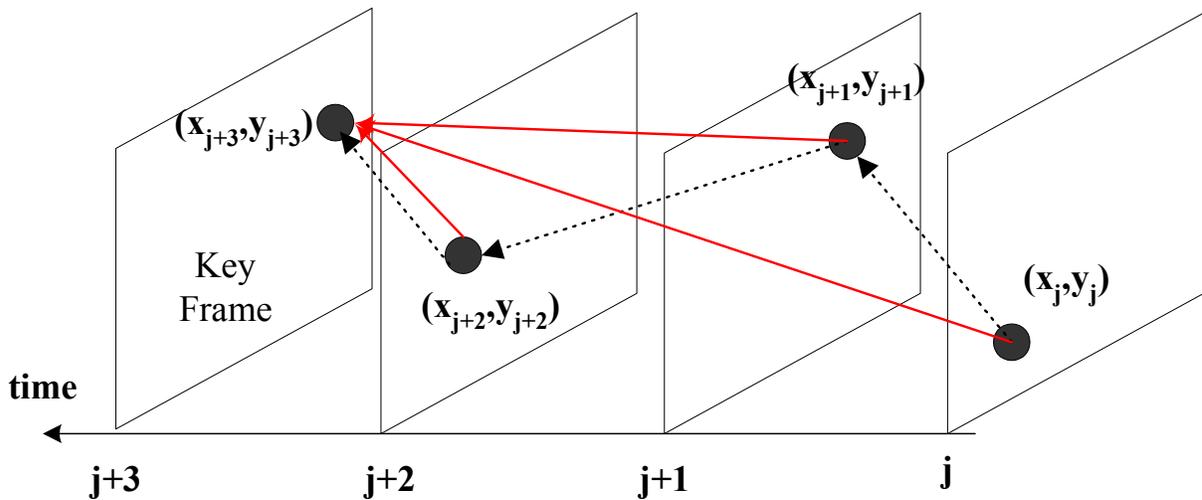


Figure 30 Intermediary Motion Vectors (Dashed Line) and Coarse Motion Vectors (Solid Line)

The common reference approach to obtain motion displacements has advantages over the 2-D motion estimation from quantized frames. Since the displacements were obtained using the original video frames, there is no distortion to the motion trajectories obtained using 2-D method. This implies that the displacements contain the actual motions of every point within the video, which is very useful information. In addition to this, the residuals are evaluated using the

original video intensities so the residuals near the key frame should approximate the intermediary model for n_x frames and have better performance than the 2-D model. The motion information obtained has the advantages over the “motion threads” model described above in that there is that the motion trajectory evaluation allows many-to-one matching rather than non-overlapping pixel-based matching, and boundary effects should also be reduced since the number of frames in the temporal dimension is the same for all the pixels.

Ideally, all the displacements obtained using this method are exact, but in practice, the displacement of any point p at time t' is very likely to increase in magnitude if t' is far away from the key frame. However, there is a constraint, that for the purpose of compression the total magnitude of all the displacement can not be no larger than those of the motion trajectories in the 2-D case. Because of this constraint, the displacements obtained for the frames not in the neighborhood of the key frame tend to be only the coarse approximation of the real displacements. By this, we will call this common reference motion estimation method the “Coarse Motion Estimation”. The residuals of each point at time i associated with this method of motion computation can be described as difference in the intensity values of the point p in the key frame and every other non-key frame, that there is a one-to-one correspondence of each residual with the original intensity values. This implies that the coarse motion compensated residual

$$r^{cr}(p) = I_i(p) - I_{key}(p), \text{ for } i \neq \text{key}$$

for any block of frames can be considered as as 3-D information, $r^{cr}(x,y,t)$, on which we can apply the 3-D multiwavelet decomposition:

$$\bar{c}_{1,1}(n,m;l) = \sum_{\rho} \tilde{H}_z(\rho - 2k) \sum_{\mu,\nu} \tilde{H}_{xy}(v - 2k, \mu - 2k) \bar{c}_0(v, \mu; \rho) = \sum_{\mu,\nu,\rho} \tilde{H}_{xyz}(v - 2k, \mu - 2k; \rho - 2k) \bar{r}^{cr}(v, \mu; \rho),$$

$$\bar{d}_{0,1;1}(n,m;l) = \sum_{\rho} \tilde{G}_z(\rho-2k) \sum_{\mu,v} \tilde{H}_{xy}(v-2k, \mu-2k) \bar{c}_0(v, \mu; \rho) = \sum_{\mu,v,\rho} \tilde{G}_{0,xyz}(v-2k, \mu-2k; \rho-2k) \bar{r}^{cr}(v, \mu; \rho),$$

$$\bar{d}_{1,1;1}(n,m;l) = \sum_{\rho} \tilde{H}_z(\rho-2k) \sum_{\mu,v} \tilde{G}_{1,xy}(v-2k, \mu-2k) \bar{c}_0(v, \mu; \rho) = \sum_{\mu,v,\rho} \tilde{G}_{1,xyz}(v-2k, \mu-2k; \rho-2k) \bar{r}^{cr}(v, \mu; \rho),$$

$$\bar{d}_{2,1;1}(n,m;l) = \sum_{\rho} \tilde{G}_z(\rho-2k) \sum_{\mu,v} \tilde{G}_{1,xy}(v-2k, \mu-2k) \bar{c}_0(v, \mu; \rho) = \sum_{\mu,v,\rho} \tilde{G}_{2,xyz}(v-2k, \mu-2k; \rho-2k) \bar{r}^{cr}(v, \mu; \rho),$$

$$\bar{d}_{3,1;1}(n,m;l) = \sum_{\rho} \tilde{H}_z(\rho-2k) \sum_{\mu,v} \tilde{G}_{2,xy}(v-2k, \mu-2k) \bar{c}_0(v, \mu; \rho) = \sum_{\mu,v,\rho} \tilde{G}_{3,xyz}(v-2k, \mu-2k; \rho-2k) \bar{r}^{cr}(v, \mu; \rho),$$

$$\bar{d}_{4,1;1}(n,m;l) = \sum_{\rho} \tilde{G}_z(\rho-2k) \sum_{\mu,v} \tilde{G}_{2,xy}(v-2k, \mu-2k) \bar{c}_0(v, \mu; \rho) = \sum_{\mu,v,\rho} \tilde{G}_{4,xyz}(v-2k, \mu-2k; \rho-2k) \bar{r}^{cr}(v, \mu; \rho),$$

$$\bar{d}_{5,1;1}(n,m;l) = \sum_{\rho} \tilde{H}_z(\rho-2k) \sum_{\mu,v} \tilde{G}_{3,xy}(v-2k, \mu-2k) \bar{c}_0(v, \mu; \rho) = \sum_{\mu,v,\rho} \tilde{G}_{5,xyz}(v-2k, \mu-2k; \rho-2k) \bar{r}^{cr}(v, \mu; \rho),$$

$$\bar{d}_{6,1;1}(n,m;l) = \sum_{\rho} \tilde{G}_z(\rho-2k) \sum_{\mu,v} \tilde{G}_{3,xy}(v-2k, \mu-2k) \bar{c}_0(v, \mu; \rho) = \sum_{\mu,v,\rho} \tilde{G}_{6,xyz}(v-2k, \mu-2k; \rho-2k) \bar{r}^{cr}(v, \mu; \rho)$$

The residuals obtained using the common reference method may be large, so there is a need to reduce the total coarse motion residuals by considering a key frame selection strategy. The purpose is to locate a key frame within a 3-D block of video frames that has a high degree of correlation with most of the other frames so as to reduce the total coarse motion residuals. Because all the video frames in the segment are all available, the selection of a key frame in 3-D coding does not have temporal constraints. A simple key frame selection strategy based on this fact will be described in the next section.

4.2 KEY FRAME SELECTION STRATEGY FOR 3-D VIDEO CODING

Our general assumption in evaluating coarse motion trajectories is that motion estimation for points in a video frame gets coarser as it is further from the key frame, so it would seem to be natural to pick the middle frame for any given block of frames as the key frame. However, in

most cases, the motion of any given point p across time may not be evenly balanced across the middle frame; it may have larger motions across certain frames than in others so the key frame should be selected to be closer to these frames in order to reduce the magnitude of the residuals. So the objective to find a key frame that produces the minimum of totals absolute residuals by evaluating every possible coarse motion trajectory from every non-key frame to the key frame. This is not realistic, so a simplified procedure is used that evaluates the sum of absolute differences in intensity values for all the points in each video frame relative to a candidate key frame, and a key frame selected as that yields the minimum sum:

$$Min_{key} \left(\sum_{(x,y)} |D_{key}(x,y)| = \sum_{(x,y)} \left| \sum_{i \neq key} I_i(x,y) - I_{key}(x,y) \right| \right)$$

where (x,y) denotes the pixel coordinates in a video frame. This procedure assumes the sum of the absolute residuals produced by a motion trajectory across two frames is proportional to the absolute intensity differences between these two frames. Under this assumption, the magnitude of motion estimation errors, or the difference in residuals, can then be translated into magnitude of the variations in intensity across time at the same position (x,y) . It then reduces the problem of find a key frame by evaluating all possible motion trajectories of all the points to the problem of locating the one frame which has the smallest total absolute difference with respect to all the other frames in the same 3-D block. However, to calculate this value for all possible frames is time consuming and unrealistic, so a logarithmic search strategy is adopted:

1. Choose frames $\frac{1}{4}$, $\frac{1}{2}$ and $\frac{3}{4}$ of the way between the first and the last frame of a given block of frames, and compute the total of the absolute difference of each of them with respect to each frame. This is equivalent to setting the middle frame as the search starting point and setting the stepsize of the search to $\pm\frac{1}{4}$ of the total number of frames around the starting point.

2. Choose the frame with the lowest total of absolute difference computed in step 1, set it as the new search starting point, and reduce the stepsize of the search to $\frac{1}{2}$ of the previous stepsize, and compute the total difference energy with respect to frames that are one stepsize away on both sides of the new starting point. Repeat this step until the stepsize is less than one.
3. Choose the frame with the lowest total energy from the last iteration of step 2, and set it as the key frame.

This greedy algorithm will only yield a key frame with a local minimum of sum of absolute differences and may not give a very small residual for encoding. However, we will demonstrate in experiments that the 3-D multiwavelet coding will compensate for it and yield good compression results even if the selected key frame is not the best possible key frame.

4.3 TWO ADDITIONAL MULTIWAVELET SYSTEMS

In addition to the two well-known multiwavelet systems considered in the last chapter, GHM and Chui-Lian, we used two additional orthogonal multiwavelet systems of multiplicity of 2, SA4 and Cardbal2, with properties similar to those of the GHM and Chui-Lian multiwavelets to investigate experimentally whether the performances of multiwavelets in low bitrate 3D video compression are better than that of comparable single wavelet.

4.3.1 Four-tapped Symmetric-Antisymmetric Orthogonal Multiwavelet (SA4)

Tham, et al [74,75] observed that it is possible to replace any given multiwavelets filter with a set of equivalent single wavelet-like filters based on its input-output characteristics, and proposed a framework of analyzing a multiwavelets filter by how well the filter pair in the equivalent set behave as low-pass and high-pass filters. Based on this, they [74,75] then introduced the 4-tapped SA4 multiwavelet which belongs to the class of symmetric-

antisymmetric multiwavelets similar to the Chui-Lian multiwavelet, but has approximation order of only 1. The following figure shows in pairs of the scaling functions and wavelets, $(\phi^1(x), \phi^2(x))$ and $(\psi^1(x), \psi^2(x))$, of the SA4 multiwavelet.

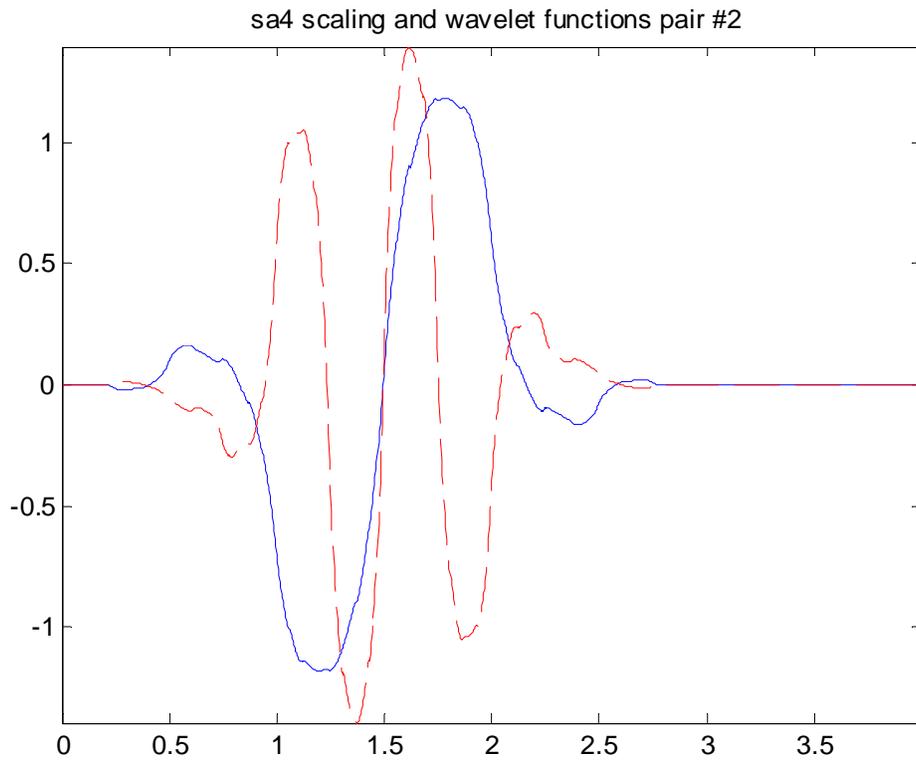
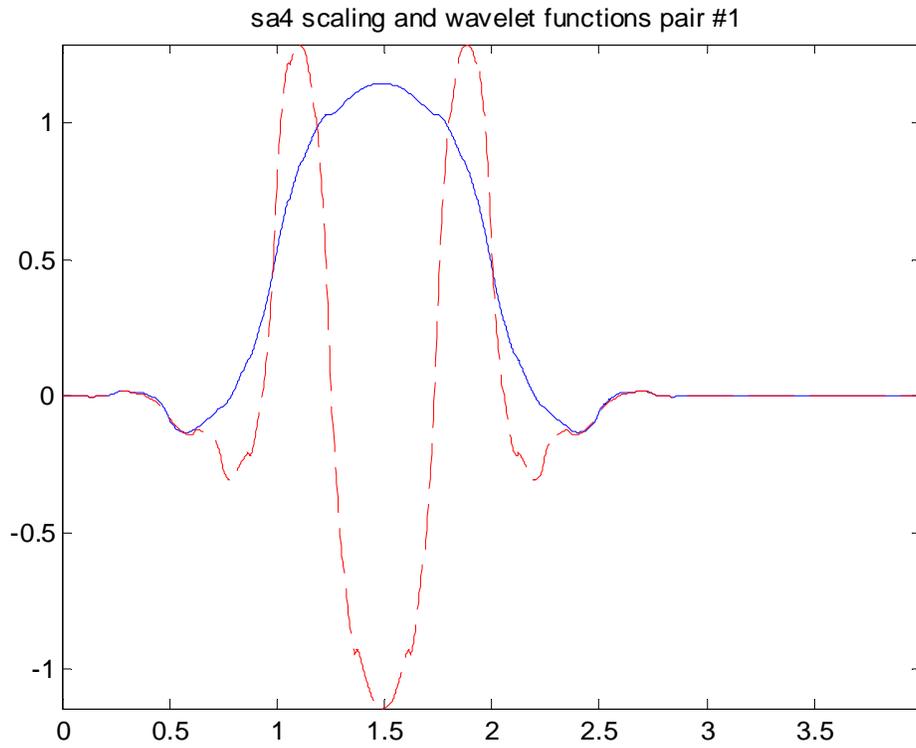


Figure 31 Scaling Functions (solid) and Wavelets (dashed) of the Symmetric-Antisymmetric Orthogonal Multiwavelet SA4 on the support $[0,4]$

The filter coefficients of $\{H_n, G_n\}$ of the SA4 multiwavelet filters are:

$$\begin{aligned}
H_0 &= \begin{bmatrix} \frac{1}{(4+\sqrt{15})^2+1} & \frac{4+\sqrt{15}}{(4+\sqrt{15})^2+1} \\ \frac{1}{(4+\sqrt{15})^2+1} & \frac{-4-\sqrt{15}}{(4+\sqrt{15})^2+1} \end{bmatrix}, & H_1 &= \begin{bmatrix} \frac{(4+\sqrt{15})^2}{(4+\sqrt{15})^2+1} & \frac{4+\sqrt{15}}{(4+\sqrt{15})^2+1} \\ \frac{-(4+\sqrt{15})^2}{(4+\sqrt{15})^2+1} & \frac{4+\sqrt{15}}{(4+\sqrt{15})^2+1} \end{bmatrix}, \\
H_2 &= \begin{bmatrix} \frac{(4+\sqrt{15})^2}{(4+\sqrt{15})^2+1} & \frac{-4-\sqrt{15}}{(4+\sqrt{15})^2+1} \\ \frac{(4+\sqrt{15})^2}{(4+\sqrt{15})^2+1} & \frac{4+\sqrt{15}}{(4+\sqrt{15})^2+1} \end{bmatrix}, & H_3 &= \begin{bmatrix} \frac{1}{(4+\sqrt{15})^2+1} & \frac{-4-\sqrt{15}}{(4+\sqrt{15})^2+1} \\ \frac{1}{(4+\sqrt{15})^2+1} & \frac{-4-\sqrt{15}}{(4+\sqrt{15})^2+1} \end{bmatrix}, \\
G_0 &= \begin{bmatrix} \frac{4+\sqrt{15}}{(4+\sqrt{15})^2+1} & \frac{1}{(4+\sqrt{15})^2+1} \\ \frac{4+\sqrt{15}}{(4+\sqrt{15})^2+1} & \frac{1}{(4+\sqrt{15})^2+1} \end{bmatrix}, & G_1 &= \begin{bmatrix} \frac{-4-\sqrt{15}}{(4+\sqrt{15})^2+1} & \frac{(4+\sqrt{15})^2}{(4+\sqrt{15})^2+1} \\ \frac{4+\sqrt{15}}{(4+\sqrt{15})^2+1} & \frac{(4+\sqrt{15})^2}{(4+\sqrt{15})^2+1} \end{bmatrix}, \\
G_2 &= \begin{bmatrix} \frac{4+\sqrt{15}}{(4+\sqrt{15})^2+1} & \frac{1}{(4+\sqrt{15})^2+1} \\ \frac{-4-\sqrt{15}}{(4+\sqrt{15})^2+1} & \frac{1}{(4+\sqrt{15})^2+1} \end{bmatrix}, & G_3 &= \begin{bmatrix} \frac{-4-\sqrt{15}}{(4+\sqrt{15})^2+1} & \frac{-(4+\sqrt{15})^2}{(4+\sqrt{15})^2+1} \\ \frac{-4-\sqrt{15}}{(4+\sqrt{15})^2+1} & \frac{(4+\sqrt{15})^2}{(4+\sqrt{15})^2+1} \end{bmatrix}
\end{aligned}$$

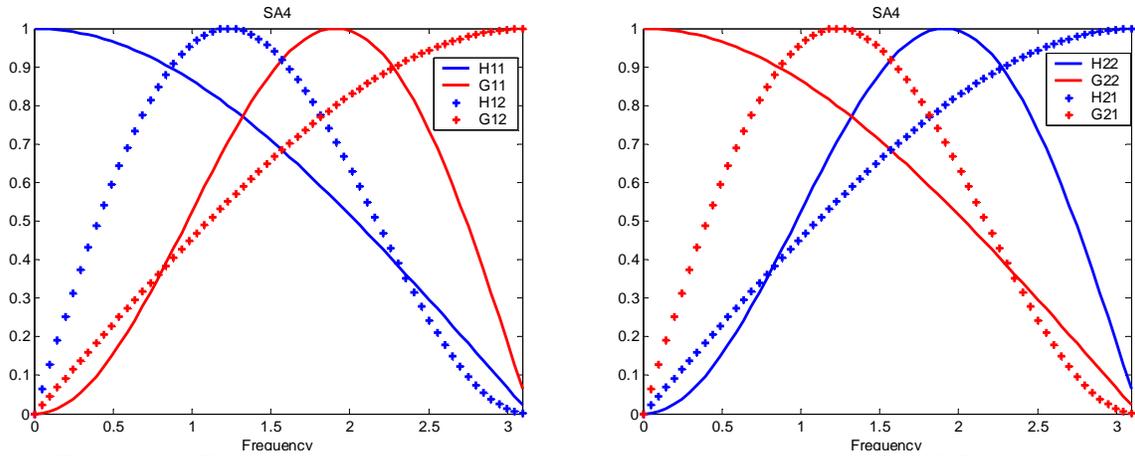
The coefficients of the approximation-based prefilter and post-filter of Strela [9] for the SA4 multiwavelet are:

$$Q = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}$$

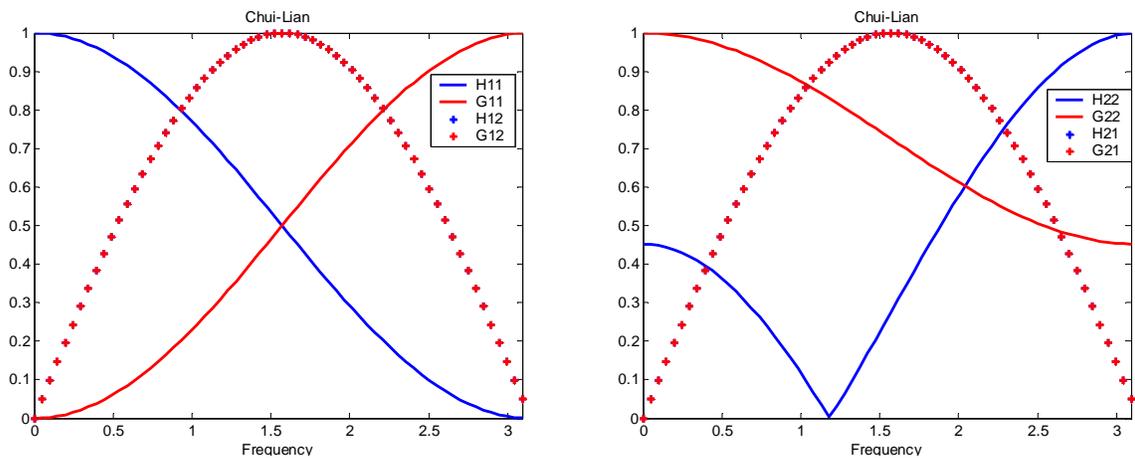
$$P = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$$

The following plots show the frequency responses of the SA4 and Chui-Lian multiwavelets. We expect these two filters will have similar performance in terms of compression due to the

properties and frequency response characteristics, such as the switching of role between H and G filters in the h_{22} , g_{22} , h_{21} , g_{21} plot, that they both share.



(a) Frequency Responses of h_{11} , g_{11} , h_{12} , g_{12} and h_{22} , g_{22} , h_{21} , g_{21} of SA4 Multiwavelet



(b) Frequency Responses of h_{11} , g_{11} , h_{12} , g_{12} and h_{22} , g_{22} , h_{21} , g_{21} of Chui-Lian Multiwavelet

Figure 32 Frequency Responses Plots of SA4 and Chui-Lian Filter Pairs (H-Blue, G-Red)

Note that $h_{ij}(\omega)$ and $g_{ij}(\omega)$ do not always exhibit the characteristics of low-pass and high-pass filters, respectively, instead they may be or band-limiting and band-pass nature in some cases, and there is a switching of roles between $g_{12}(\omega)$ and $h_{21}(\omega)$ or $h_{22}(\omega)$. This reflects unbalances of the two scaling bases. Otherwise, the characteristics of these two symmetric-antisymmetric multiwavelets are similar. This leads us to expect that the performances of SA4 and Chui-Lian multiwavelets will be similar in terms of compressions.

4.3.2 Two-Balanced Cardinal Orthogonal Multiwavelet (Cardbal2)

If the two scaling bases have similar frequency characteristics, the multiwavelets are called balanced multiwavelets. The balanced multiwavelet bases carry over their zero moment properties to their corresponding discrete filters. As mentioned in Chapter 2, one of the functions of the approximation-based prefiltering is to generate a more balanced 2-row ($r=2$) vector from a 1-dimensional signal of balanced multiwavelets, this balancing property is already there and hence no prefiltering is required. A scaling function $\phi(t)$ is an interpolatory or cardinal function if $\phi(n)$ satisfies

$$\phi(n) = \delta(n) = \begin{cases} 1, & \text{for } n = 0 \\ 0, & \text{for } n = \pm 1, \pm 2, \dots \end{cases}$$

$$\text{where } \phi(x) = \sum_n \phi(n) \text{sinc}(x-n), \quad \text{sinc}(x) = \frac{\sin(\pi x)}{\pi x}$$

A multiwavelet with interpolatory scaling functions $\phi(x)$ is called a cardinal multiwavelet. Selesnick [76,77,78] used the requirements (constraints) of orthogonality, cardinal scaling functions and compact support to generate a set of orthogonal balanced cardinal multiwavelets of various approximation orders. We considered the two-balanced cardinal multiwavelet that possesses the approximation order of two, same as the GHM and Chui-Lian multiwavelets, although it has a longer support $[0,5]$ and filter length of 6. Figure 33 shows the scaling and wavelet functions of the two-balanced cardinal multiwavelet, hereto referred as “Cardbal2”. Selesnick generated Cardbal2 in such a way that its second scaling function $\phi^2(x)$ is a shifted (by $\frac{1}{2}$) version of its first scaling function $\phi^1(x)$, so the two scaling bases are balanced and the 1-D input signal can be converted to a vector input of two components by simply taking even-

indexed samples and odd-indexed samples in the original signal, respectively. The 2-scale matrix dilation is given by

$$\phi(x) = \sqrt{2} \sum H(n) \phi(2x - n) \quad \text{where } H(n) = \begin{bmatrix} h_0(2n) & h_0(2n+1) \\ h_1(2n) & h_1(2n+1) \end{bmatrix}$$

$$\text{Similarly, } \psi(x) = \sqrt{2} \sum G(n) \phi(2x - n) \quad \text{where } G(n) = \begin{bmatrix} g_0(2n) & g_0(2n+1) \\ g_1(2n) & g_1(2n+1) \end{bmatrix}.$$

The elements of $\{h_0(k)\}$ and $\{h_1(k)\}$ are filter coefficients two filters for the two scaling bases.

Similar, $\{g_0(k)\}$ and $\{g_1(k)\}$ filter coefficients two filters for the two wavelets.

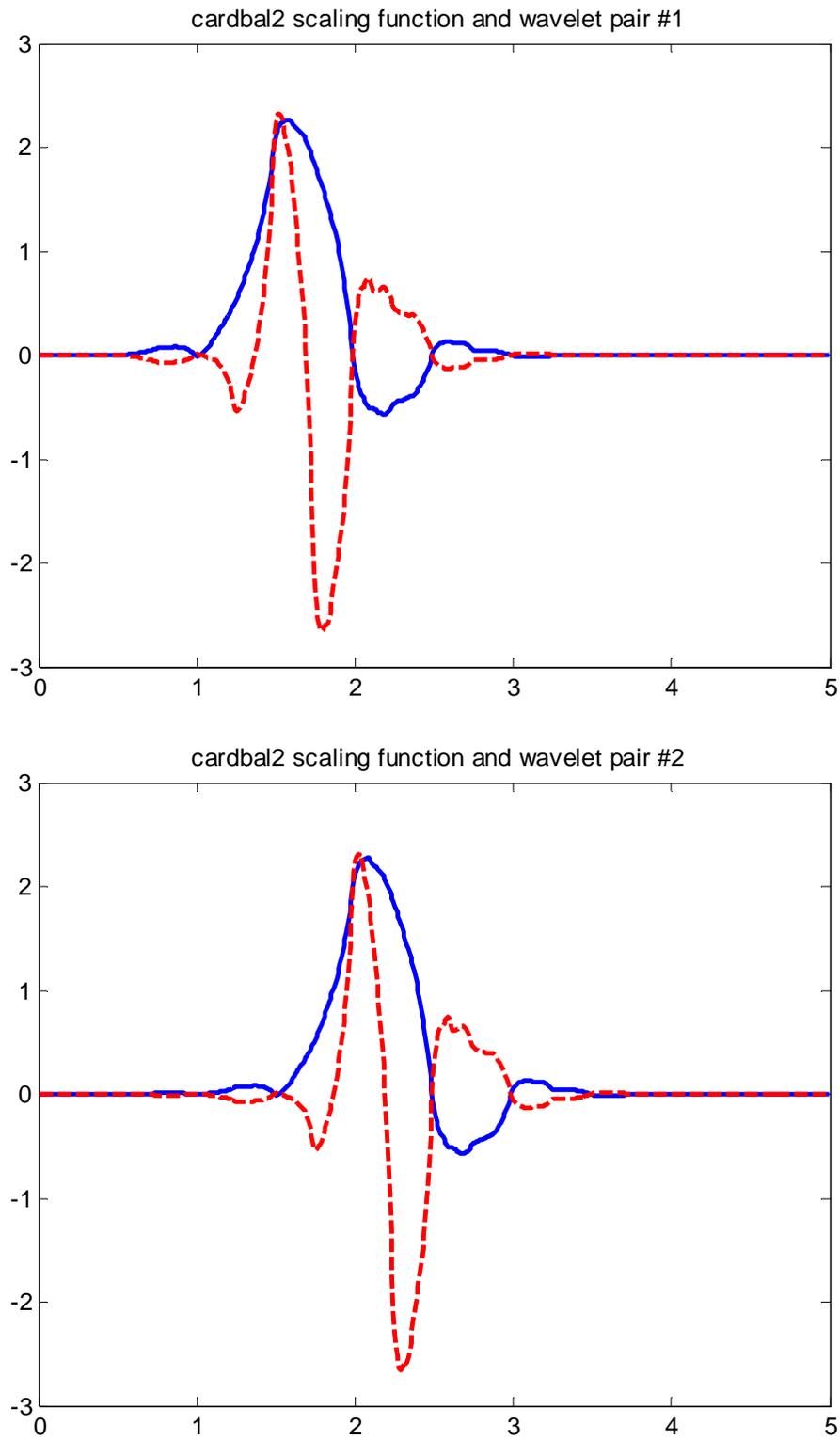


Figure 33 Two-Balanced Cardinal Multiwavelet Cardbal2 on the support $[0,5]$; Scaling Functions are shown in Blue/Solid Line, and Wavelets shown in Red/Dashed Line

The filter coefficients H_n , and G_n (for $n=0,1,2,..5$) of the Cardbal2 multiwavelet are given by:

$$H(0) = \begin{bmatrix} \frac{1}{32} & 0 \\ \frac{4-\sqrt{15}}{32} & 0 \end{bmatrix}, \quad H(1) = \begin{bmatrix} \frac{4+\sqrt{15}}{32} & 1 \\ \frac{15}{16} & 0 \end{bmatrix}, \quad H(2) = \begin{bmatrix} \frac{15}{16} & 0 \\ \frac{8-\sqrt{15}}{16} & 1 \end{bmatrix},$$

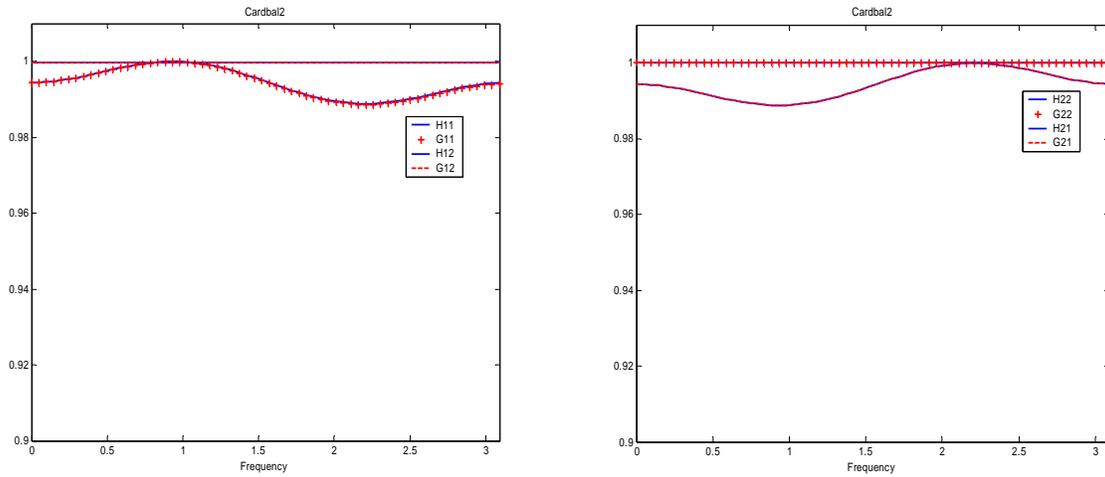
$$H(3) = \begin{bmatrix} \frac{\sqrt{15}-8}{16} & 0 \\ \frac{15}{16} & 0 \end{bmatrix}, \quad H(4) = \begin{bmatrix} \frac{15}{16} & 0 \\ \frac{-4-\sqrt{15}}{32} & 0 \end{bmatrix}, \quad H(5) = \begin{bmatrix} \frac{\sqrt{15}-4}{32} & 0 \\ -\frac{1}{32} & 0 \end{bmatrix}$$

and

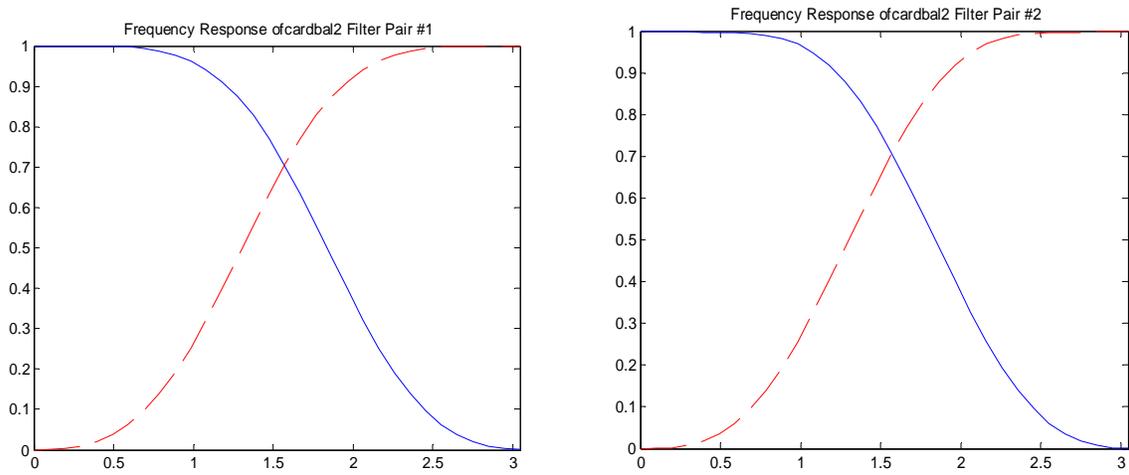
$$G(0) = \begin{bmatrix} -\frac{1}{32} & 0 \\ \frac{\sqrt{15}-4}{32} & 0 \end{bmatrix}, \quad G(1) = \begin{bmatrix} \frac{-4-\sqrt{15}}{32} & 1 \\ -\frac{15}{16} & 0 \end{bmatrix}, \quad G(2) = \begin{bmatrix} -\frac{15}{16} & 0 \\ \frac{\sqrt{15}-8}{16} & 1 \end{bmatrix},$$

$$G(3) = \begin{bmatrix} \frac{8-\sqrt{15}}{16} & 0 \\ \frac{15}{16} & 0 \end{bmatrix}, \quad G(4) = \begin{bmatrix} -\frac{15}{16} & 0 \\ \frac{4+\sqrt{15}}{32} & 0 \end{bmatrix}, \quad G(5) = \begin{bmatrix} \frac{4-\sqrt{15}}{32} & 0 \\ \frac{1}{32} & 0 \end{bmatrix}$$

The frequency responses of $\{h_{ij}(w)\}$ and $\{g_{ij}(w)\}$ of the matrix filter components are shown in Figure 34(a), and the frequency responses of $H_0(w)$, $H_1(w)$, $G_0(w)$, and $G_1(w)$, are shown in Figure 34(b), the latter conforms to the characteristics of low-pass and high-pass filters, respectively.



(a) Frequency Responses of h_{11} , g_{11} , h_{12} , g_{12} and h_{22} , g_{22} , h_{21} , g_{21} Filters



(b) Frequency Responses of h_0 , g_0 and h_1 , g_1 Filters (H-solid, G-dashed)

Figure 34 Frequency Responses Plots of Cardbal2 Filters

Although the Cardbal2 multiwavelet has longer support and filter length in comparison to the GHM and Chui-Lian multiwavelets, its balancing properties makes it a better multiwavelet for some applications. Cardbal2 is currently the balanced cardinal multiwavelet with the shortest support and filter length. It does not require prefiltering to obtain balanced input rows, which may be obtained by simply splitting the input into a row of even-indexed samples and another row of odd-indexed samples.

4.4 MODIFIED 3-D SPIHT ENCODER FOR COARSE MOTION RESIDUALS

One characteristic of coarse motion residuals is that the maximum bit required to code each frame vary temporally; the maximum bit required to code the wavelet coefficients of those frames near the key frame should be lower than that of those frames farther away from the key frame. This is not a problem for the 2D SPIHT, where the maximum bit for each frame can be different from each other, however, in the 3D SPIHT, the initial lists of LIP and LIS must contain all the coefficients in the LLL subband and be tested at the same threshold even if the maximum bit for every frame is different, thus resulting in wasted bits in the case like coarse motion residuals.

The modified 3D SPIHT algorithm is as follows:

0. Initialize three linked lists:

LIP: List of Insignificant Pixels (wavelet coefficients)

LIS: List of Insignificant Sets

(The insignificant set associated with an insignificant pixel or coefficient in LIP is a sub-tree originating from the significant pixel. A tree may be of type D which include all children and all generations below or it could be of type L which includes all grandchildren and below but not the immediate children)

LSP: List of Significant Pixels (pixels above current testing threshold)

1. Determine the maximum bit required to code the absolute value of all coefficients in each temporal frame (and all of its children, grandchildren and below) in the coarsest subband and send them. Set 'n' as the number of bits required representing the coefficient with the largest absolute value in binary form.

2. Extract the pixels (coefficients) from the each frame in root node (LLL) that require n bits to code and insert them in LIP and insert the trees associated with them into LIS as type D. Set LSP to empty set. Do not insert frames requiring fewer than n bits to code.

3. Set the threshold as $2^{n-1}-1$ so that all the coefficients whose absolute values require n binary bits to represent will be above the threshold and be significant.

4. Sorting pass:

III. Check all LIP elements against the threshold, and move all the significant elements from LIP to LSP after sending a '1' and its sign bit. A '0' is sent for each insignificant element in the LIP.

IV. Check all LIS elements for significance.

- For a 'D' type tree, send a '1' if one of the children or grandchildren and below is significant, then check each child against the threshold. If a child is significant, output a '1' plus its sign and move it to LSP set; but if a child is not significant, output a '0' and move it to the LIP. If grandchildren exist, move this tree to the end of LIS as type L, otherwise remove it from LIS. A '0' is sent if the entire D tree is insignificant.
- For an 'L' type tree, send a '1' if one of the grandchildren and below is significant, remove the parent from LIS add each child to the end of LIS as type 'D', otherwise send a '0'.

5. Refinement pass:

- Check each of the original elements in the LSP, which are elements in LSP before the sorting pass, against the threshold and output a '1' if significant, and a '0' if otherwise.

6. Reduce the testing threshold, 'n', by 1 bit, and if $n > 0$, then extract the pixels (coefficients) from each frame in the root node (LLL) that require n bits to code, insert them in the beginning of LIP and insert the trees associated with them into the beginning of LIS as type D. Do not insert frames requiring less than n bits to code. Repeat step 3, etc.

This modified SPIHT algorithm is designed to save some '0' bits by not testing coefficients in frames that are known to be below the threshold. This will increase the efficiency of coding the coarse motion residuals. This design can be translated to 2-D SPIHT in the cases where the large coefficients in each frame are concentrated in certain spatial regions or more bits are required for foveated areas in each frame.

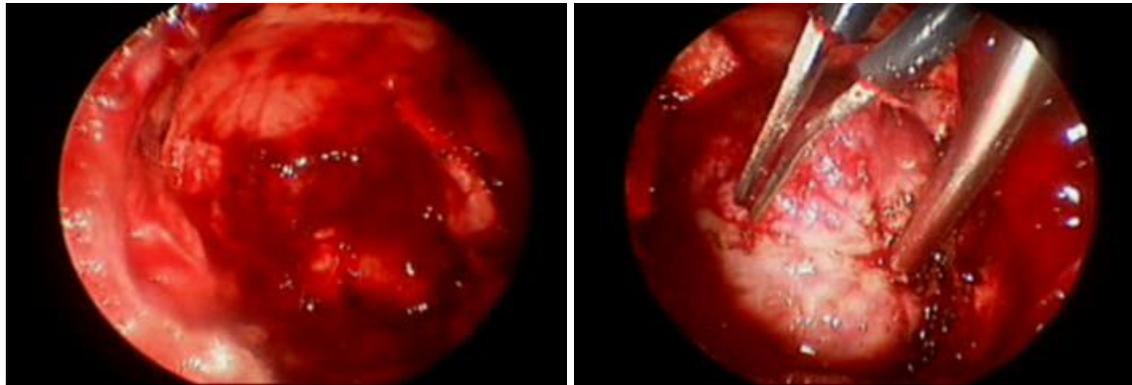
4.5 EXPERIMENTS AND RESULTS

This experiment was designed to test the performance of 3D multiwavelet-based coarse motion residuals coding system, with and without the modified SPIHT coder using segments of medical endoscopy video sequences as input.

There were two reasons for using the endoscopy video sequences; (1) Some chromatic information are almost the same through each video sequence, we reduced the experimental parameters by performing experiments in the luminance domain only and then translating the result into color domain, (2) the frames were acquired through a traversal in a contiguous 3-D environment that lends itself naturally to the paradigm of the 3-D based video coding.

The three endoscopy video sequences (with frame size 240x352) were titled M,P, and R. From M, P and R, total of nineteen video segments of thirty three frames each were extracted at random to be used as test data. Eleven segments (M01 to M11) were from sequence M which is the longest, and four each from sequence P (P01 to P04) and sequence R (R01 to R04).

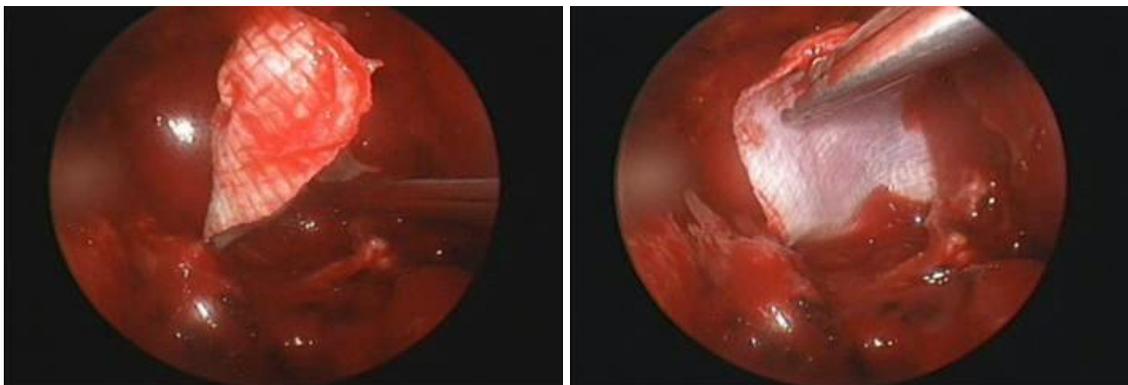
Snapshots of these videos are shown in Figure 35. In each segment, one frame was chosen as key frame.



(a) Snapshots from the M endoscopy video sequence



(b) Snapshots from the P endoscopy video sequence



(c) Snapshots from the R endoscopy video sequence

Figure 35 Snapshots from Three Neurosurgical Endoscopy Video Sequences

For calculating motion trajectories, the full-search block-based motion estimation was used, where the residuals for pixels in the same rectangular block were evaluated with every trajectory up to $\pm p$ in both x and y directions, as indicated in Figure 36..

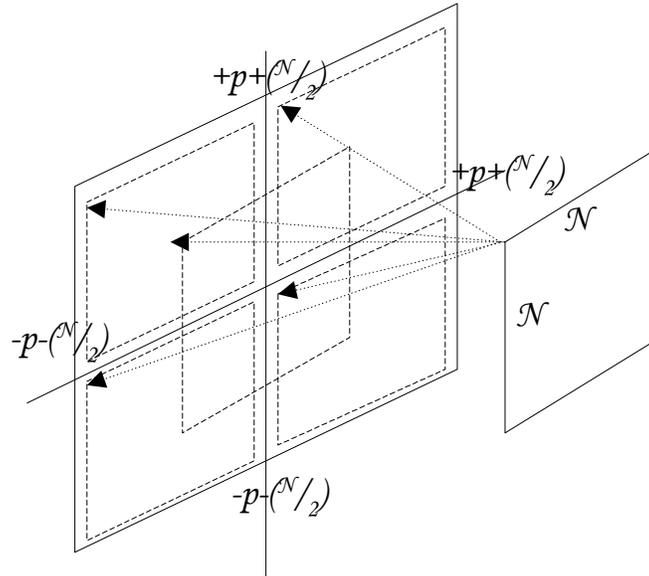


Figure 36 Search for the Best Match of the $N \times N$ Block within a Larger Block

The full-search block-based motion estimation algorithm was based on the six-level computation for Minimum Absolute Difference (MAD) as stated below.

```

Do h=1 to Nh
Do v=1 to Nv
Dmin(h, v)=∞
MV(h, v)=(0,0)
Do m=-p to p
Do n=-p to p
    MAD(m, n)=0
    Do j=1 to N
    Do i=1 to N
        MAD(m, n)=MAD(m, n) + (1/√2)X((h-1)N+i, (v-1)N+j) - Y((h-1)N+i+m, (v-1)N+j+n) /
    End Do i, j
    If ( MAD(m, n) < Dmin(h, v)
        Dmin(h, v) = MAD(m, n)
        MV(h, v) = (m, n)
    End If
End Do n,m,h,v

```

h, v – horizontal and vertical coordinates of a block
i, j – coordinates of a pixel within a block
m, n – horizontal and vertical coordinates of a vector
MV – motion vector, *p* – search range, *N* – block size in horizontal and vertical direction

Figure 37 The Six-loop Full-Search Block Matching Algorithm

As can be seen from the algorithm, the MAD values were computed by comparing the residuals generated by matching any given block of size $N \times N$ in frame X with blocks in reference frame Y within the range of $(2p+1) \times (2p+1)$ in order to determine the most like displacement of the block in X to Y . In the experiment, both N and p were set to be 16 (4 bits), i.e. the size of each block was 16×16 and the maximum possible displacement was ± 16 . Three multiwavelets were used in the experiment: the Chui-Lian multiwavelet, the SA4 multiwavelet and the Cardbal2 multiwavelet. Approximation prefiltering was put forth for the the Chui-Lian multiwavelet and also for the SA4 multiwavelet. The Cardbal2 multiwavelet, being a balanced interpolative multiwavelet requires no prefiltering and 2-channel input in each dimension was obtained by splitting the data into the even-indexed and odd-indexed samples (for $r=2$). For comparison

purpose, DB4 was chosen. Only 2-levels of decomposition was used in all cases. Experiments were carried out for the 3-D coarse motion residual coding using unmodified and modified 3D SPIHT coders as well as 2D multiwavelet and single wavelet coder for comparison purposes. In term of coding, the 3D multiwavelet coefficients are coded using the unmodified 3D symmetric SPIHT while the 2D multiwavelet and single-wavelet coefficients are coded using regular 2D SPIHT. The setting for the SPIHT quantizer in both 2D and 3D cases is 0.2bpp or 1/40 compression ratio. In the experiments, the following parameters and results are taken or measured: the key frame selected for each test video sequence, total sum of residuals, and the averaged PSNR measurements of reconstructed frames.

Table 1 lists the averaged PSNR over 32 frames for the 19 endoscopy segments. It is noted that 3D multiwavelet coarse motion residuals video coding outperforms the 2D cases in all 19 cases. We underscored the best averaged PSNR value for each segment, and noted that no 2D system, using either the multiwavelet Chui-Lian or the single wavelet DB4 had gain over the 3D systems even once; i.e. the averaged PSNR for every 3D multiwavelet coding is higher than the 2D methods. For example, the 3D Chui-Lian had gains over the 2D Chui-Lian, from $\sim 0.4\text{dB}$ (M10) to $\sim 5\text{dB}$ (P04), while the 2D single wavelet has the worst performance in most cases. The winner for each segment is almost evenly distributed among the three multiwavelets with no clear overall winner as the differences in their performances are small. This may be due to the fact that the level of decomposition was limited to only two, and the contributions from different bases could not be clearly distinguished. As expected, the SA4 multiwavelet, though only provides the approximation order of 1, has performance similar to Chui-Lian. This shows that the characteristics of the filters are at least as significant as approximation order in the performance of low bitrate 3D wavelet-based compression using multiwavelets. It is noted that

the best sequence for the multiwavelet Chui-Lian is different in 3-D case (M03) than that of 2-D (R01), it may imply that the results of 3-D multiwavelet-based compression can not be directly inferred from those of 2-D even if the same multiwavelet is used, but the sample space is too small to state this conclusively.

Table 1 Average PSNR over 32 Endoscopy Video Frames at 0.2bpp

Sequence	3D Cardbal2	3D SA4	3D Chui-Lian	2D DB4	2D Chui-Lian
M01	<u>35.75(best)</u>	35.44	35.62	33.97	34.77
M02	36.96	<u>37.38(best)</u>	36.50	33.53	34.31
M03	37.54	<u>37.72 (best)</u>	37.02	33.70	34.55
M04	28.68	<u>29.87(best)</u>	28.74	27.23	27.63
M05	<u>30.27(best)</u>	29.69	30.19	28.54	28.61
M06	<u>30.27(best)</u>	29.69	30.19	28.49	28.96
M07	35.06	35.49	<u>35.62(best)</u>	31.72	32.73
M08	26.93	<u>27.57(best)</u>	27.31	26.3	26.53
M09	27.75	27.76	<u>27.92(best)</u>	27.03	27.56
M10	26.91	26.92	<u>27.07(best)</u>	26.22	26.77
M11	<u>27.74(best)</u>	27.69	27.15	26.46	26.82
R01	<u>38.00 (best)</u>	37.61	36.66	34.63	35.11
R02	37.20	<u>37.40(best)</u>	36.92	34.51	35.01
R03	32.94	<u>33.42(best)</u>	33.11	30.64	30.81
R04	<u>32.40(best)</u>	32.27	31.44	30.19	30.11
P01	34.20	<u>34.42(best)</u>	34.16	32.38	33.06
P02	32.09	<u>32.27(best)</u>	32.11	29.99	30.46
P03	36.16	<u>36.95(best)</u>	36.58	33.09	33.65
P04	<u>36.20(best)</u>	35.77	35.96	30.93	31.08

The plots in Figures 38, 39 and 40 represent the PSNR measurements of sequences M03, R02, and R03, respectively. The '+' symbols are the results of 3-D multiwavelet (CL), the solid line is the result of 2-D multiwavelet (CL) and the dashed line is the result of 2-D single wavelet (DB4). These are the cases where 3D CL had clear advantages over the 2D cases in terms of averaged PSNR. As expected, those frames nearest to the key frame have better performance in all cases. Even in the cases where the averaged performance improvement was not as significant, such as M05 shown in Figure 41, the stability of 3D coding over 2D coding is evident because the 2D coding showed sharp drop in PSNR values when two consecutive frames had low correlation, while the 3D coding result maintains its shape.

It should be noted that in rare cases, such as M08 shown in Figure 42, the 3D Chui-Lian has lower performance in 1/3 of the frames even though it still maintained a higher averaged PSNR. However, this is a difference that the modified SPIHT can help overcome, as discussed below.

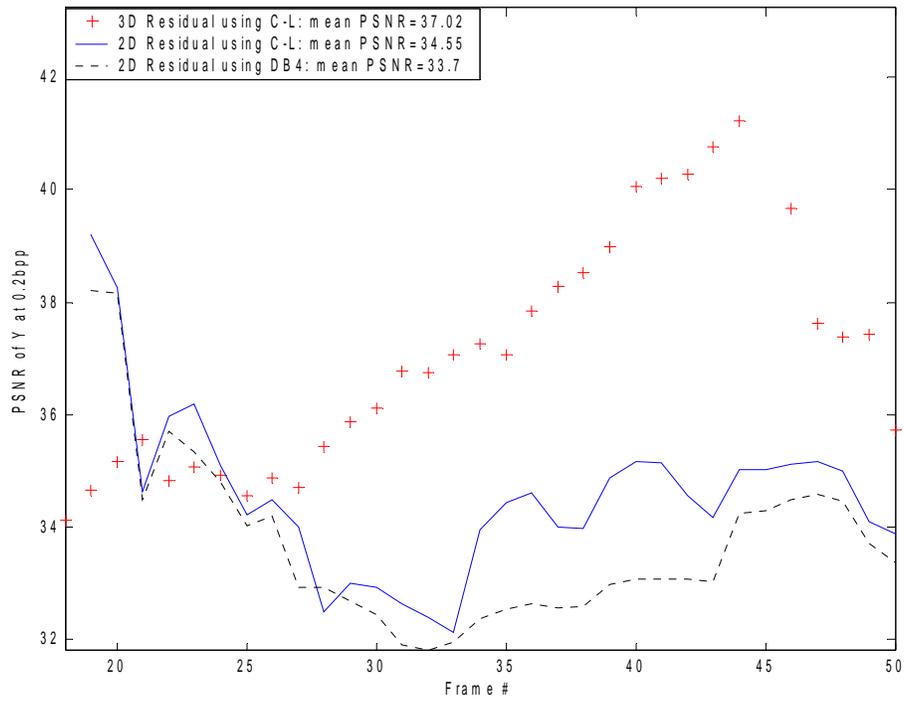


Figure 38 PSNR Measure of 32 Video Frames of Endoscopy sequence M03

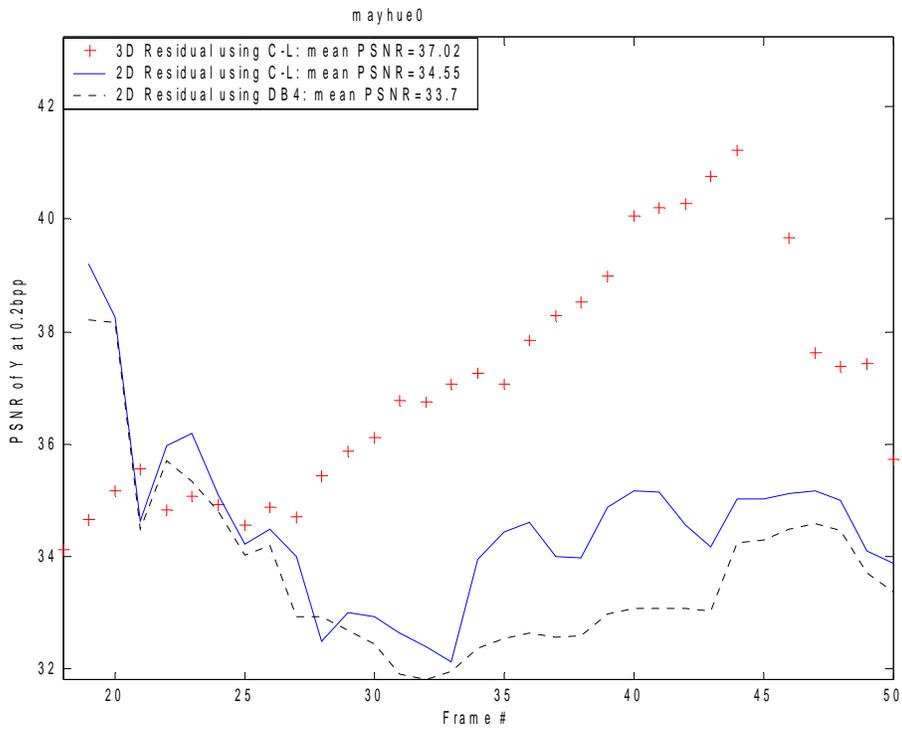


Figure 39 PSNR Measure of 32 Video Frames of Endoscopy sequence R02

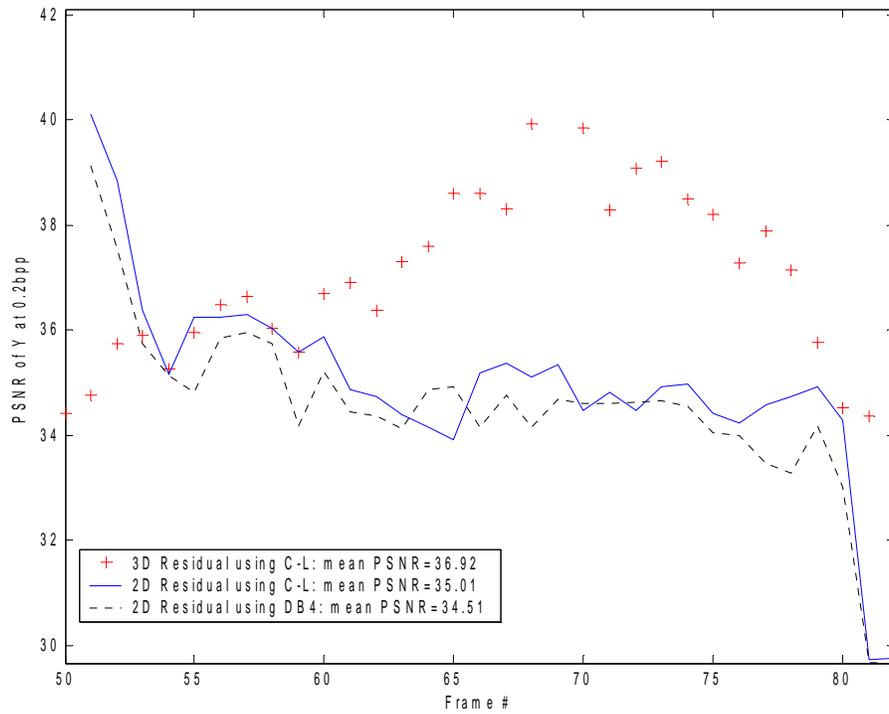


Figure 40 PSNR Measure of 32 Video Frames of Endoscopy sequence R03

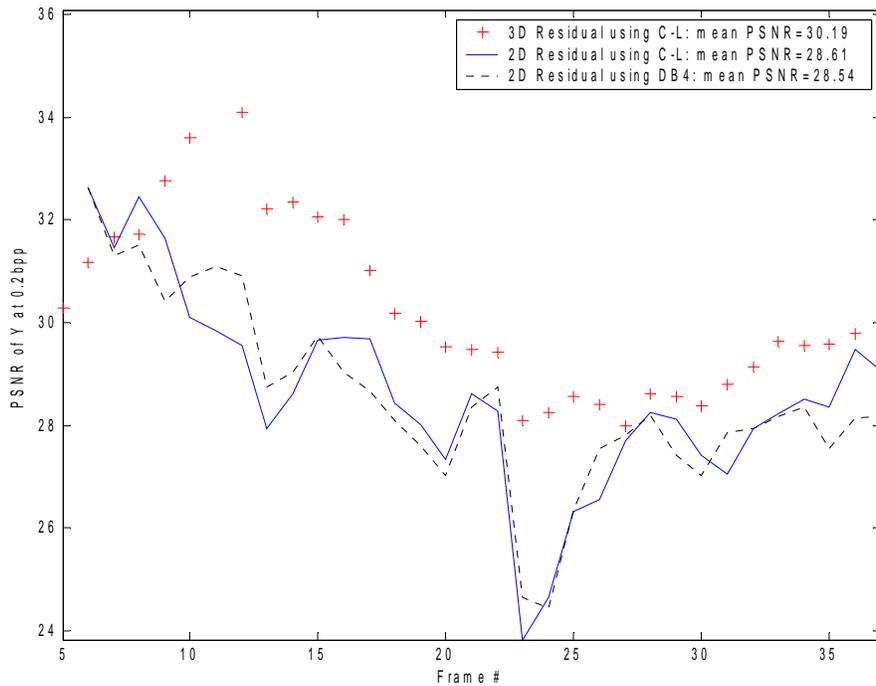


Figure 41 PSNR Measures of Endoscopy sequence M05

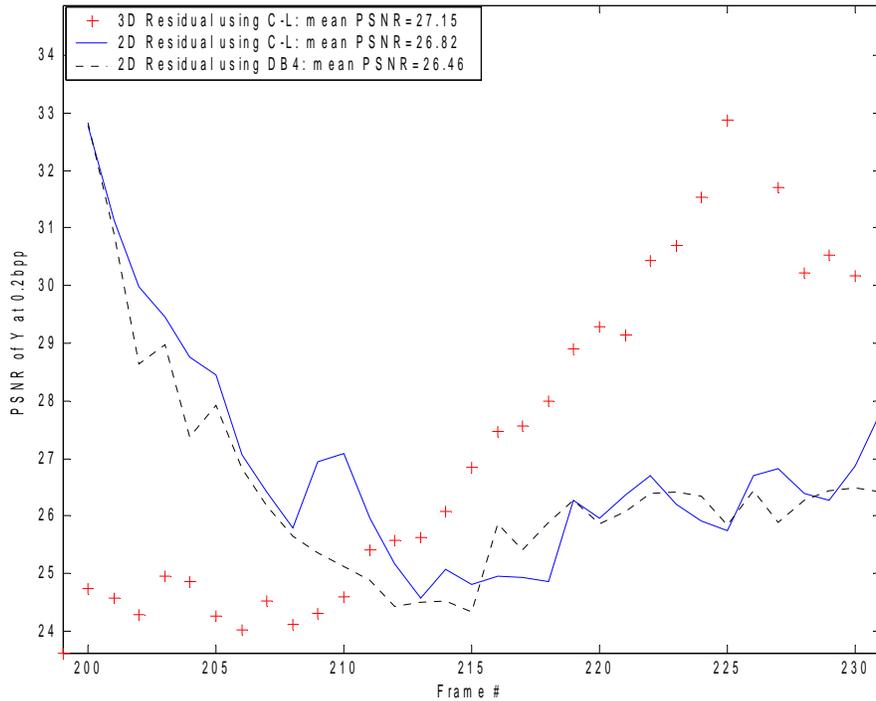


Figure 42 PSNR Measures of Endoscopy sequence M08

In a second experiment designed to test the efficacy of the modified 3-D SPIHT coder, we repeated the experiment above for the 3D video coder with just the Chui-Lian multiwavelet, but replaced the regular 3D SPIHT coder with the modified version. Four out of the nineteen neurosurgical segments whose performance in the first experiment needed improvement, M08, M09, M10 and M11, they were chosen as the testing sequences here. For the purpose of performance comparison, the results of the 3-D multiwavelet-based compression of the same frames using their original intensity values were also presented, and Table 2 shows the results. As it can be seen, the unmodified SPIHT helped in improving the PSNR in all four cases by as much as 1dB at the low bitrate of 0.2bpp. The results of using the 3D multiwavelet coding described in Chapter 3 on the original intensity values using the unmodified SPIHT is given in

the first column of Table 2 for comparison purposes. The 3D coarse motion residuals plus the modified 3D SPIHT encoder showed performance gains over the direct 3D multiwavelet coding by as much as 3.5dB.

Table 2 Average PSNR over 32 Endoscopy Video Frames at 0.2bpp

Sequence	Regular 3D Multiwavelet coding using regular 3D SPIHT	3D Coarse Motion Residuals coding using regular 3D SPIHT	3D Coarse Motion Residuals coding using modified 3D SPIHT
M08	24.31	27.31	27.87
M09	25.20	27.92	28.44
M10	24.94	27.07	27.50
M11	24.49	27.15	28.17

4.6 SUMMARY

It was postulated at the beginning that the compression performance of multiwavelet-based 3-D video coding would be superior to the performance of 2-D wavelet coding using comparable wavelets, but it was difficult to evaluate and compare these two different methods on an even platform. We have developed the “coarse-motion-residual” approach for 3D video coding as a platform for the comparison. The “coarse-motion-residual” approach utilizes a common reference frame extracted from a video segment as a key frame. A key frame selection strategy was developed, with the goal of reducing total residuals, based on the concept that the larger the motions are, the greater the residuals will be which are computed by the difference between the intensity values of two frames. A modified version of 3-D SPIHT is also developed, which is more efficient by taking advantage of the spatio-temporal relationship of the “coarse-motion-residuals”.

Three orthogonal multiwavelets, Chui-Lian's CL, Tham's SA4 and Selesnick's Cardbal2, were used in the experiment which was performed for comparing the multiwavelet-based 3-D coarse-motion-residuals coding against 2-D motion residuals coding using single wavelet and multiwavelet. The key frame selection strategy reduces the average of total prediction error from a possible 29.604×10^6 per frame to a value of 16.848×10^6 using multiwavelet CL, which is more comparable to the average of 13.825×10^6 obtained by the 2-D video coding using wavelet DB4. But even with the high average residual per frame to code, the 3-D multiwavelet "coarse-motion-residuals" coding outperformed the 2-D motion residuals video coding in the experiment by 0.3dB to 4.88dB(multiwavelet CL), 0.2dB to 4.69dB(multiwavelet SA4), and 0.19dB to 5.12dB (multiwavelet Cardbal2), even at the low bit-rate of 0.2bpp. The simulation results showed that the best sequence for Cardbal2, whose matrix filters' frequency response exhibits characteristics of low-pass and high-pass filters, is the same sequence for the 2-D cases; yet the best sequence Chui-Lian, whose matrix filters' frequency response do not resemble that of 2-D, is different in the case of 3-D than 2-D. The direct interpretation is that 3-D multiwavelet may be better at exploiting the differences between the contributions from different basis functions than 2-D multiwavelet in compression, but the sample space is too small to state this conclusively. The use of the modified SPIHT with multiwavelet CL in four sequences (that are more difficult to compress) enhanced the performance by 0.5dB to 1.0dB at 0.2bpp. This experiment not only demonstrated the general superiority of 3-D multiwavelet over the standard wavelet video coding, but has also revealed the advantage of 3-D multiwavelet coding over 2-D multiwavelet even having to deal with more residuals without using the modified SPIHT coder. The same experiment showed that the incorporation of motion information into the 3-D multiwavelet coder can improve compression performance by 2 to 3dB using the four sequences.

Although we expected the key frame selection strategy to have good performance, the results shown in Table 3 implies that a better strategy may be desired. Several points are noted from the Table 3: First three columns in the table, it is seen that the key frames selected by the proposed algorithm are generally closer to the global minimum in terms of total residuals, shown in the “Best Key” and “Worst Key” columns, respectively; the “Worst Key” refers to the case that resulted in global maximum total residuals. These numbers were obtained by exhaustively searching through all possible key frame using displacements of each frame with respect to them. In some cases, such as M07, P02, P03 and P04, the key frame selection algorithm actually found the global minima. It is also noted that although in a number of cases the total residuals of the coarse motion estimation resulted from the greedy key frame selection algorithm were higher than those obtained by the nearest-neighbor motion estimation, they were actually smaller in some instances; e.g. R01, P03 and P04. An improved algorithm may be required. In the next chapter, we will explore the effects of using multiwavelet packets in 3D video coding.

Table 3 Totaltotal Residuals summed over 32 Endoscopy Video Frames

Total Sum of Block Motion Residuals of 32 Video Frames (pre-quantized) x10 ⁶					
3D Coarse Motion Residuals (Chui-Lian)				2D (Nearest-Neighbor)	
Sequence	Best Key	Worst Key	Selected Key	DB4	Chui-Lian
M01	9.8069	13.175	9.9955	7.7977	7.4534
M02	8.8156	11.276	9.0202	8.1574	7.8897
M03	7.8626	11.276	8.0167	8.1200	7.5883
M04	51.175	213.51	52.458	36.611	35.741
M05	16.396	21.217	17.287	15.043	14.922
M06	20.454	15.532	17.436	15.096	14.853
M07	9.8154	14.521	9.8154	10.232	9.5958
M08	29.216	40.272	29.578	21.549	21.134
M09	26.253	40.441	26.461	20.142	19.179
M10	30.258	40.923	31.446	22.189	21.174
M11	28.842	37.471	29.498	21.695	20.739
R01	6.6513	9.5675	6.6519	6.9875	6.8259
R02	6.9125	11.885	6.9550	7.0954	6.8788
R03	11.989	20.411	12.405	10.765	10.441
R04	11.897	17.900	15.975	11.473	11.460
P01	9.0247	13.283	9.4586	9.0778	8.6685
P02	12.661	17.515	12.661	12.130	11.619
P03	7.3502	10.851	7.3502	8.3740	7.9062
P04	7.4143	11.254	7.4143	10.108	9.9391

5.0 MULTIWAVELET PACKETS IN 3-D COARSE MOTION RESIDUALS CODING

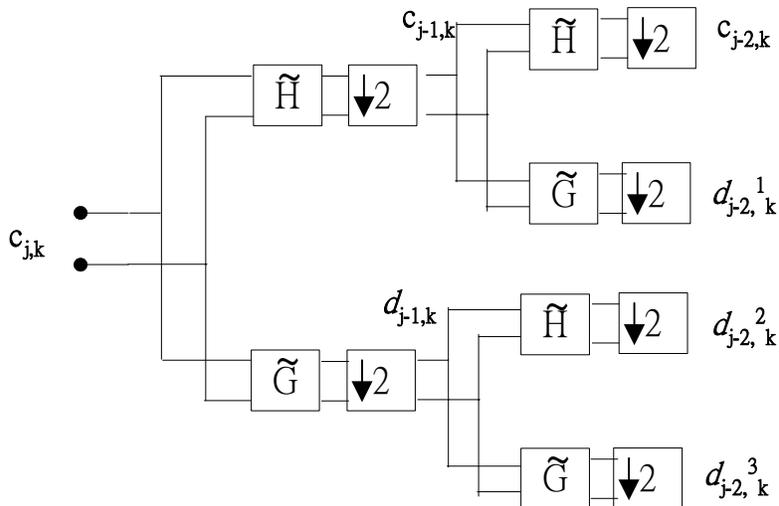


Figure 43 Standard 1-D Multiwavelet ($r=2$) Packet Decomposition for Two Levels

The wavelet packet transform is similar to the wavelet transform except that not only the scaling coefficients at each scale are decomposed, the wavelet coefficients are also decomposed; the decomposition follows a balanced binary decomposition tree, as shown in Figure 43 for a 1-D multiwavelet ($r=2$) decomposition for two scale levels. As indicated in Figure 43, $c_{0,k}$ is the prefiltered 2-channel ($r=2$) input vector sequence that is decomposed into the scaling component $c_{-1,k}$ and the wavelet component $d_{-1,k}$, both components can be further decomposed following a balanced decomposition tree. At a decomposition level J , there will be 2^J packets, and the basis function (vector function) $\psi^{(q)}(x)$ in the q^{th} packet, ($q=0,1,\dots,2^{J-1}$), are derived from the scaling function ϕ and wavelet ψ ;

$$\Psi^{(0)}(x) = \phi(x), \quad \Psi^{(1)}(x) = \psi(x)$$

$$\Psi^{(2n)} = \sqrt{2} \sum_k \tilde{H}(k) \Psi^{(n)}(2x - k)$$

$$\Psi^{(2n+1)} = \sqrt{2} \sum_k \tilde{G}(k) \Psi^{(n)}(2x - k)$$

The multiwavelet packet decomposition is then given by:

$$\bar{c}_{j-2,n} = \sum_m \tilde{H}_{m-2n} \bar{c}_{j-1,m} = \sum_m \sum_k \tilde{H}_{m-2n} \tilde{H}_{k-2m} \bar{c}_{j,2k}$$

$$\bar{d}_{j-2,n}^1 = \sum_m \tilde{G}_{m-2n} \bar{c}_{j-1,m} = \sum_m \sum_k \tilde{G}_{m-2n} \tilde{H}_{k-2m} \bar{c}_{j,k}$$

$$\bar{d}_{j-2,n}^2 = \sum_m \tilde{H}_{m-2n} \bar{d}_{j-1,m} = \sum_m \sum_k \tilde{H}_{m-2n} \tilde{G}_{k-2m} \bar{d}_{j,k}$$

$$\bar{d}_{j-1,k}^3 = \sum_m \tilde{G}_{m-2n} \bar{d}_{j-1,m} = \sum_m \sum_k \tilde{G}_{m-2n} \tilde{G}_{k-2m} \bar{d}_{j,k}$$

These equations of decompositions are more complex than they appear. Let us first examine the frequency content of $c_{j-1,k}$ and $d_{j-1,k}$ in Figure 44.

$$\bar{c}_{j-1,k}(2\omega) = \begin{bmatrix} \tilde{H}_{11}(\omega) & \tilde{H}_{12}(\omega) \\ \tilde{H}_{21}(\omega) & \tilde{H}_{22}(\omega) \end{bmatrix} \bar{c}(\omega), \quad \bar{d}_{j-1,k}(2\omega) = \begin{bmatrix} \tilde{G}_{11}(\omega) & \tilde{G}_{12}(\omega) \\ \tilde{G}_{21}(\omega) & \tilde{G}_{22}(\omega) \end{bmatrix} \bar{c}(\omega),$$

The frequency function of these filters, $\{\tilde{H}_{ij}(\omega)\}$ and $\{\tilde{G}_{ij}(\omega)\}$, are shown in Figure 43, where filtering of $(\tilde{H}_{11}, \tilde{H}_{21}, \tilde{G}_{11}, \tilde{G}_{21})$, is induced by the first basis function and that of $(\tilde{H}_{12}, \tilde{H}_{22}, \tilde{G}_{12}, \tilde{G}_{22})$ by the second basis function.

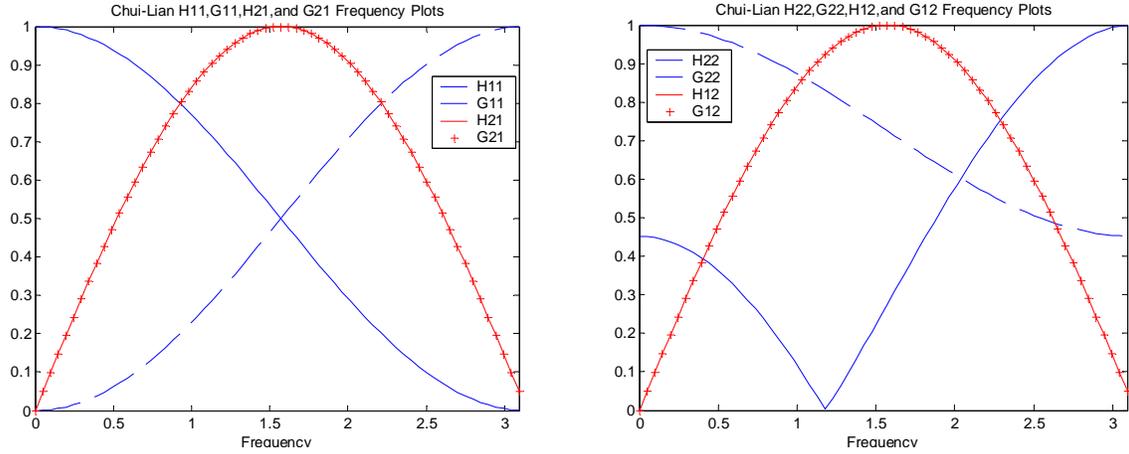


Figure 44 Frequency Responses of the Chui-Lian Multiwavelet Filters $\{\tilde{H}_{ij}(\omega)\}$ and $\{\tilde{G}_{ij}(\omega)\}$

Indeed, the $\tilde{H}_{11}(\omega)$ and $\tilde{G}_{11}(\omega)$ functions behave like low-pass and high-pass filters respectively while $\tilde{H}_{21}(\omega)$, and $\tilde{G}_{21}(\omega)$ behave like overlapping band-pass filters in the domain of the first basis function. It is interesting to note that $\tilde{G}_{22}(\omega)$ and $\tilde{H}_{22}(\omega)$ behave differently and their frequency characteristics appear to be reversed. For the four packets at the level 2, the Chui-Lian (CL) packet basis functions $(\psi_1^{(k)}(x), \psi_2^{(k)}(x))$, $(k=0, 1, 2, 3)$, are shown in Figure 45, their respective frequency spectra $(\psi_1^{(k)}(\omega), \psi_2^{(k)}(\omega))$ are shown in Figure 46. It shows that the frequency coverages of these four packets are in the increasing order as they appear naturally, in the Chui-Lian multiwavelet, without any switching as with wavelet packets in the single wavelet system.

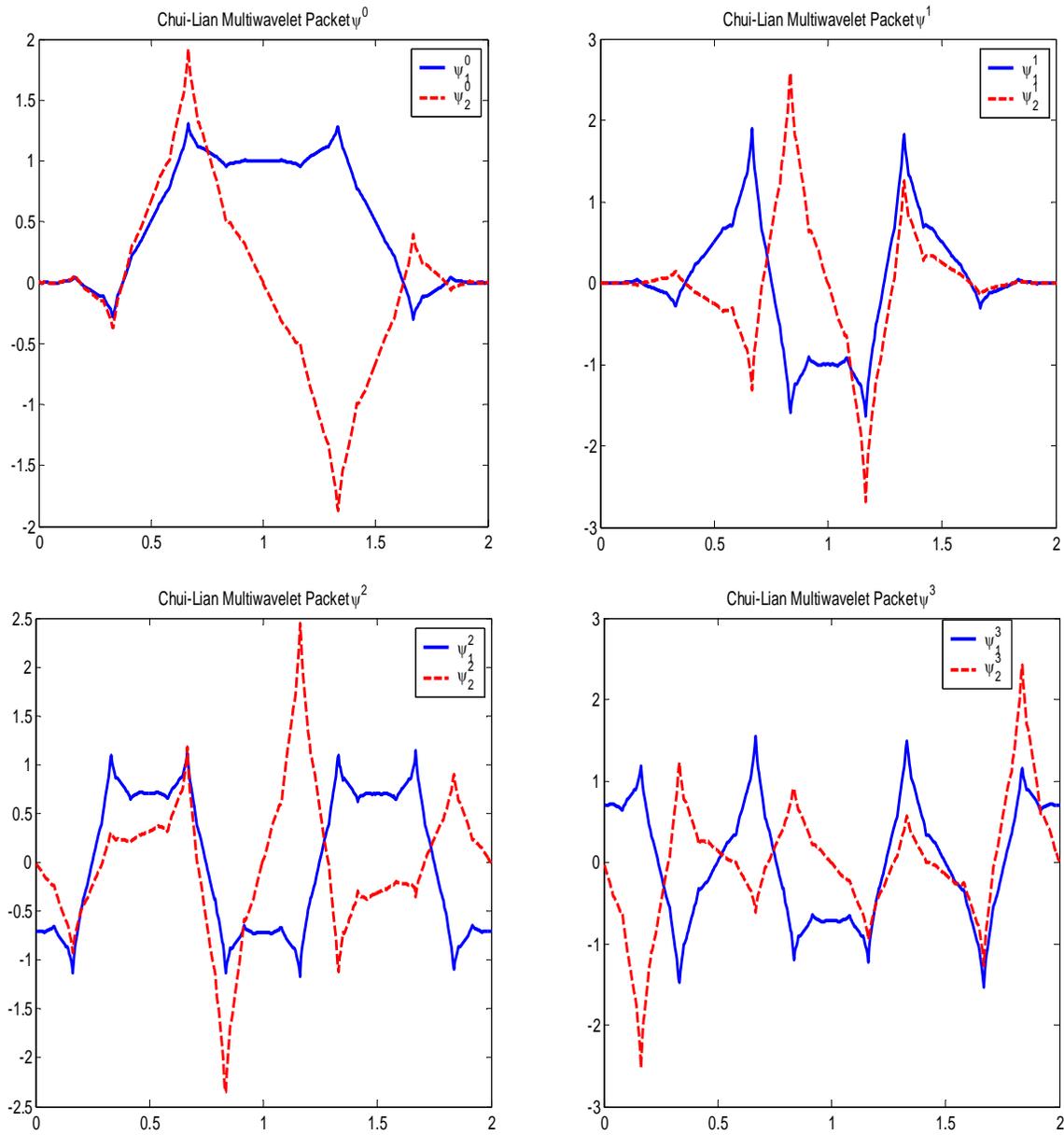


Figure 45 ($\psi_1(k)(x)$, $\psi_2(k)(x)$), ($k=0,1,2,3$) of Chui-Lian Multiwavelet Packets at Scale Level 2

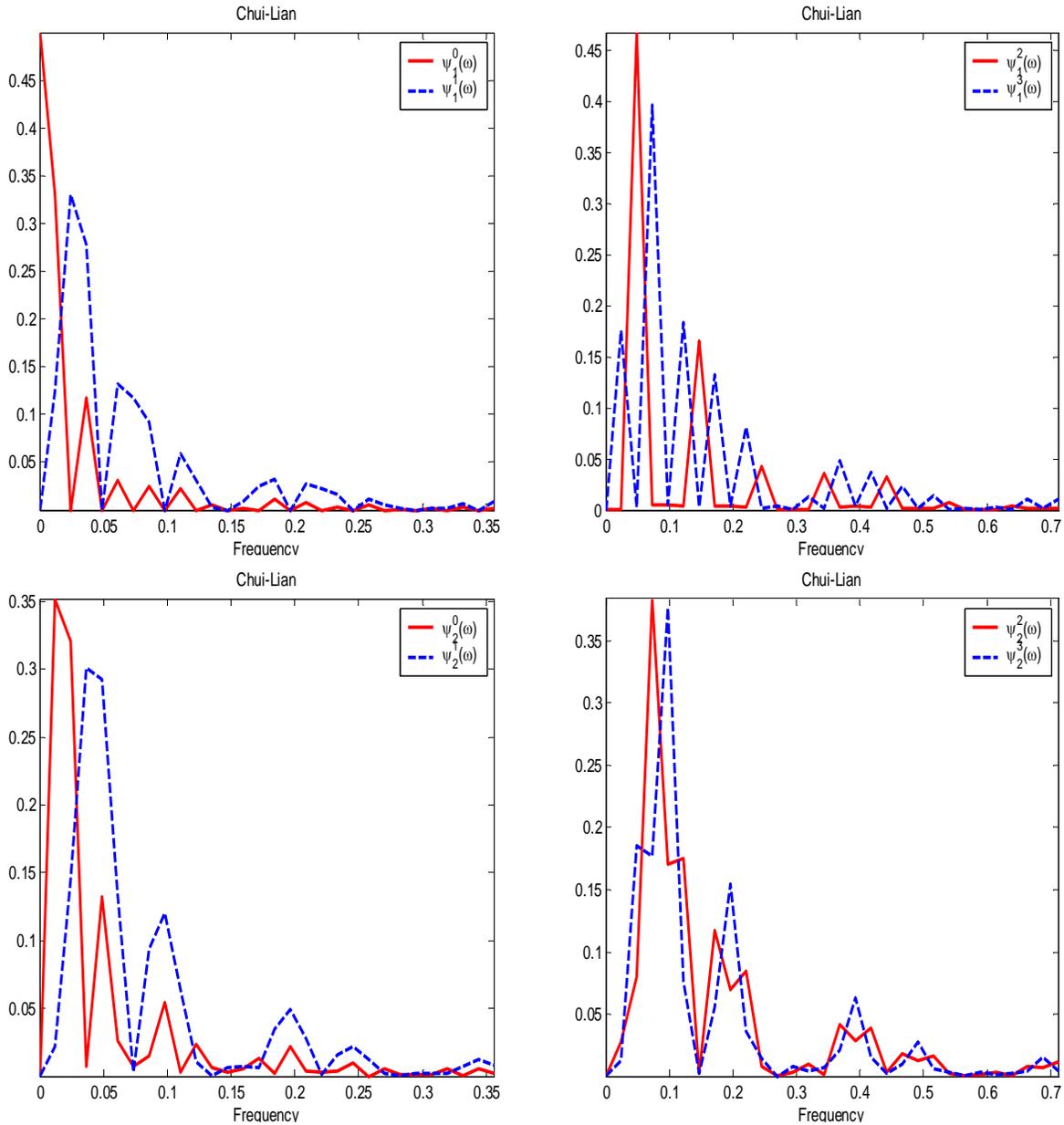


Figure 46 The Frequency Spectra of the $\psi(k)(\omega)$ ($k=0,1,2,3$) of the Chui-Lian Multiwavelet

It is speculated that (multi)wavelet packet transform would not have as much contribution at higher compression ratios than at low ratios because many natural images contain significantly more low frequency than high frequency information. However, in the case of coarse motion residuals, there is as much information of interest at higher frequencies as at lower frequencies that leads to the motivation to explore whether complete-tree multiwavelet packet decomposition

can be may to improve the multiwavelet-based 3-D coarse motion residuals coding under low bit-rate situations.

5.1 WAVELET PACKETS SPIHT STRUCTURE

Since the wavelet packet decomposition decomposes the data to packets at a given scale, there is no longer a clear parent-children relationship which exists across scales in the SPIHT for the regular wavelet decomposition. When we attempt to use SPIHT structure as the platform of encoding packet coefficients, the parent-children relationships need to be redefined or modified to suit the wavelet packet decomposition; in this case, the packets would appear to be siblings which different closeness in relationships. Since all the packets subbands in a complete-tree decomposition are equal in size, a natural way to adapt the SPIHT structure for wavelet packets is to assign each parent pixel only one “child” pixel in each of the “close” packets, but each parent node will still retain the same number of children nodes (now cousin nodes) as in the regular SPIHT structure so it will not change the basic SPIHT algorithm. The new relationships are shown below in Figure 47 for the case of 2-D:

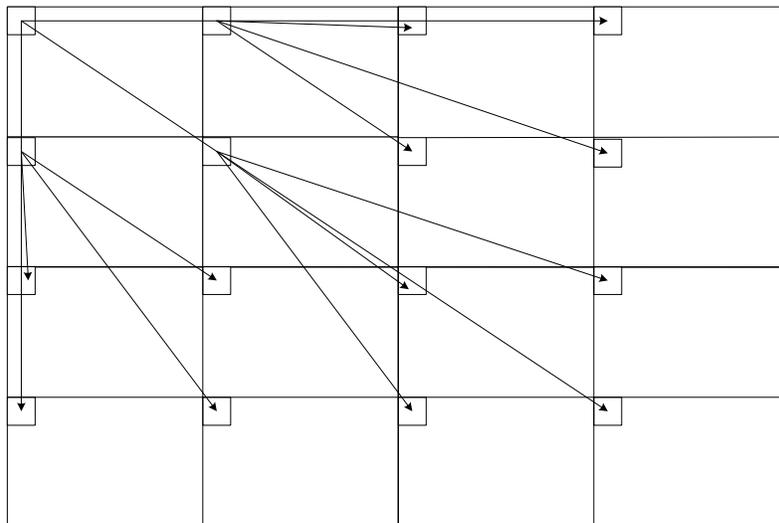


Figure 47 A Modified “Parent-Children” Relationship for SPIHT in Wavelet Packet Decomposition at Level 2

It is noted that since the complete-tree wavelet packet decomposition decomposes a given signal into equal-sized subbands, the plausibility of shuffling or re-ordering the subbands based on certain criterion may exist. So in addition to the regular 3-D multiwavelet packet, we will also explore the effect of shuffling based on two different criteria on video compression.

5.2 PACKET ORDERING BY FREQUENCY

In multiwavelet packet transform, there are r basis functions in each wavelet packet ($\Psi^{(k)}(x) = (\psi_1^{(k)}(x), \dots, \psi_r^{(k)}(x))$). We will examine the frequency coverage of each packet as contributed by the spectra of r basis functions. Let us consider, in particular, the Chui-Lian multiwavelet CL ($r=2$) and its four wavelet packets at scale level 2, where the low-pass and high pass filter matrices $H(k)=[h_{ij}(k)]$ and $G(k)=[g_{ij}(k)]$ are given by

$$\{h_{11}(k)\} = \left\{ \frac{1}{4} \frac{1}{2} \frac{1}{4} \right\}, \{h_{12}(k)\} = \left\{ \frac{1}{4} \ 0 \ -\frac{1}{4} \right\}, \{h_{21}(k)\} = \left\{ -\frac{\sqrt{7}}{8} \ 0 \ -\frac{\sqrt{7}}{8} \right\}, \{h_{22}(k)\} = \left\{ -\frac{\sqrt{7}}{8} \ \frac{1}{4} \ -\frac{\sqrt{7}}{8} \right\}$$

and

$$\{g_{11}(k)\} = \left\{ -\frac{1}{4} \ \frac{1}{2} \ -\frac{1}{4} \right\}, \{g_{12}(k)\} = \left\{ -\frac{1}{4} \ 0 \ \frac{1}{4} \right\}, \{g_{21}(k)\} = \left\{ \frac{1}{8} \ 0 \ -\frac{1}{8} \right\}, \{g_{22}(k)\} = \left\{ \frac{1}{8} \ \frac{\sqrt{7}}{4} \ \frac{1}{8} \right\}$$

Without the loss of generality, let us use the frequency responses of \tilde{H}_{11} and \tilde{G}_{11} filters of multiwavelet CL to demonstrate this phenomenon.

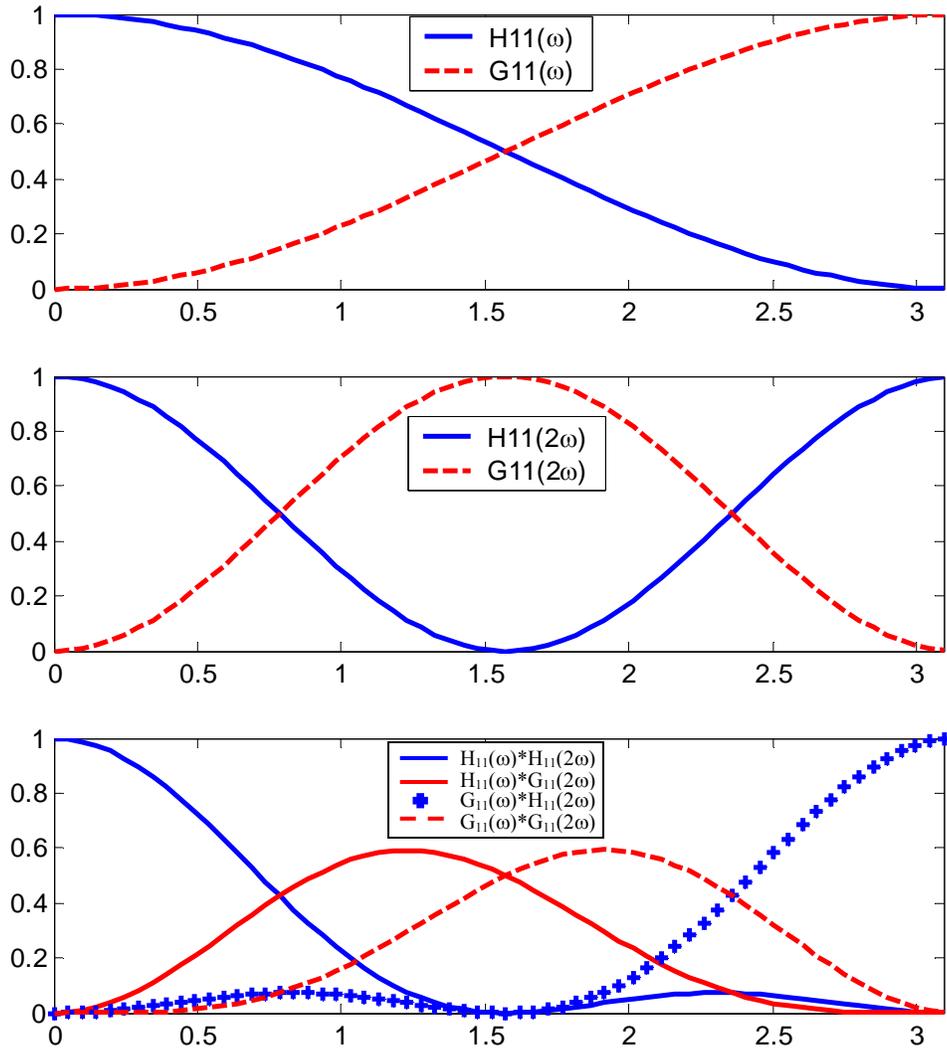


Figure 48 Frequency Coverages of $G_{11}(\omega)$, $H_{11}(\omega)$ and $G_{11}(2\omega)$, $H_{11}(2\omega)$ in the Situation of One Wavelet Basis Function

Let us consider the first basis function alone, $\psi_1^{(1)}(x)$, which is equivalent to considering a single wavelet case, and the associated filter functions, \tilde{H}_{11} and \tilde{G}_{11} . Where decomposed a signal for two levels using wavelet packet decomposition, the first and second packets are generated by the low-pass filter \tilde{H}_{11} in the first level decomposition followed, after down-sampling, by \tilde{H}_{11} and \tilde{G}_{11} filters in the second level, respectively. The third and fourth packets

are generated by the high-pass filter \tilde{G}_{11} in the first level followed, after down-sampling, by the \tilde{H}_{11} and \tilde{G}_{11} filters. The first two packets would contain the frequencies determined by $\tilde{H}_{11}(\omega)$ $\tilde{H}_{11}(2\omega)$ and $\tilde{H}_{11}(\omega)$ $\tilde{G}_{11}(2\omega)$, and have the frequency orders of 1 and 2 respectively. The third packet contains the frequencies determined by $\tilde{G}_{11}(\omega)$ $\tilde{H}_{11}(2\omega)$ that be in a higher frequency range, as shown in the bottom plot of Figure 48, than $\tilde{G}_{11}(\omega)$ $\tilde{G}_{11}(2\omega)$ that determines the frequency coverage of the fourth packet; so the frequency order of the 3rd and 4th packets are actually 4 and 3, respectively. But this is based on the argument of one wavelet only.

Base on this observation, a simple method can be used to generate the frequency order for wavelet packets for any number of decomposition levels by the use of the ‘binary labeling method’:

1. For decomposition level 1: label the packet generated by the H filter as ‘0’ and the packet generated by the G filter as ‘1’.
2. For any subsequent level, the label of each packet inherits its parent’s binary label and is appended by either a ‘0’ or ‘1’ depending on the type of its parent. If the parent of this packet is generated by using G filter, and if this packet is generated by using H filter, then a ‘1’ is appended to the end of its label, otherwise a ‘0’ is appended. Or, if the parent of this packet is generated by using H filter, and this packet is also generated by using H filter, then a ‘0’ is appended to its label, otherwise a ‘1’.
3. Convert the binary label of any packet to its decimal equivalence which is the frequency order of that packet at the given decomposition level.

This ordering process for three levels of wavelet packet decomposition of 1-D signals is illustrated in Figure 49. It naturally provides a switching of frequency order as necessary.

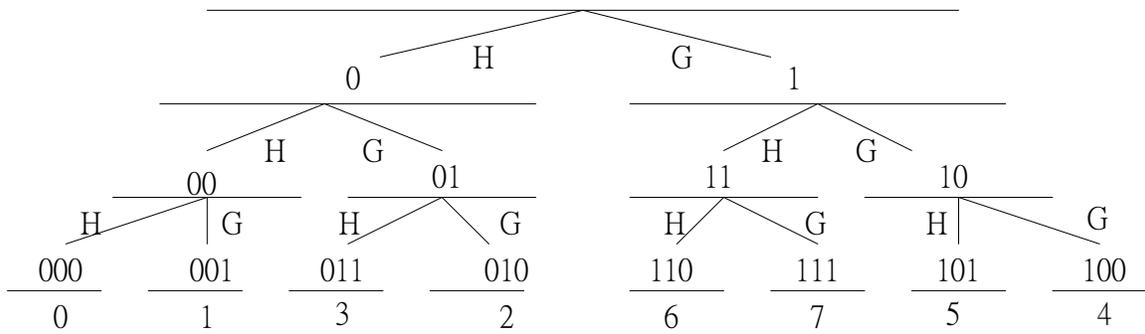


Figure 49 Frequency Ordering of 1D Wavelet Packets for up to Three Levels of Decomposition in the Single Wavelet System

Thus, for single wavelet systems, this method of determining frequency order of any packet can be easily translated to 2D and 3D by considering each dimension separately. For two level decomposition of a 2D image, the vertical packets will each receive a ‘0’ as its vertical label if H filter is used in the vertical decomposition, and a ‘1’ if G filter is used; vice versa for the horizontal packets. Let us use the notation “(vertical label, horizontal label)” to represent the frequency order of each packet, then the LL packet would have (0,0) as its label, the HH packet would have (1,1) as its label, and HL and LH will have (0,1) and (1,0) as their labels, respectively. The rule for binary labeling in either horizontal or vertical direction should be applied separately for each successive level of decomposition along either direction, as shown in Figure 50. For 3-D wavelet packets, an extra binary label will be used to form a triplet representation. Note that this frequency ordering does not require extra overhead cost in wavelet packet-based video compression, since both sender and receiver can determine the frequency order of each packet independently without the need for extra information.

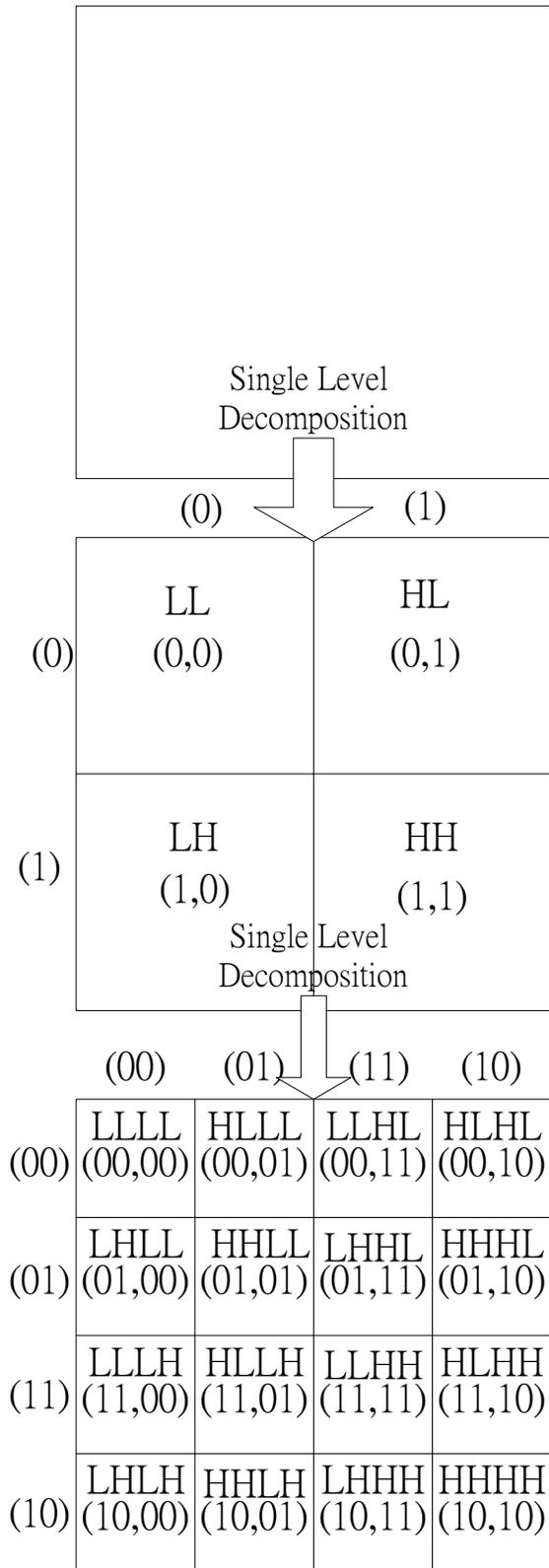


Figure 50 Frequency Ordering of 2D Wavelet Packets for up to Two Levels of Decomposition in the Single Wavelet System

This is illustrated by the frequency response of the 2D filters for the first basis function of the CL multiwavelet. Figures 51, 52 and 53 respectively show the frequency responses of the filters $\tilde{H}_{11}(\omega_y)\tilde{H}_{11}(\omega_x)$, $\tilde{G}_{11}(\omega_y)\tilde{G}_{11}(\omega_x)$, and $\tilde{G}_{11}(\omega_y)\tilde{H}_{11}(\omega_x)$ with $\tilde{H}_{11}(\omega_y)\tilde{G}_{11}(\omega_x)$.

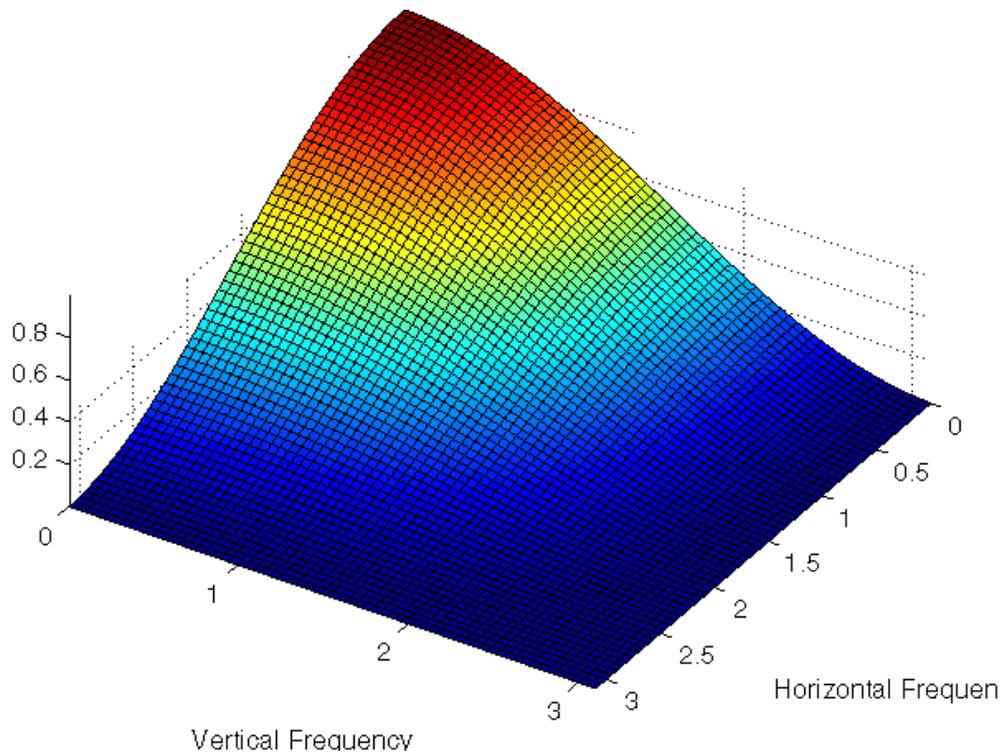


Figure 51 Frequency Response of $H_{11}(\omega_y)H_{11}(\omega_x)$ for Chui-Lian Multiwavelet

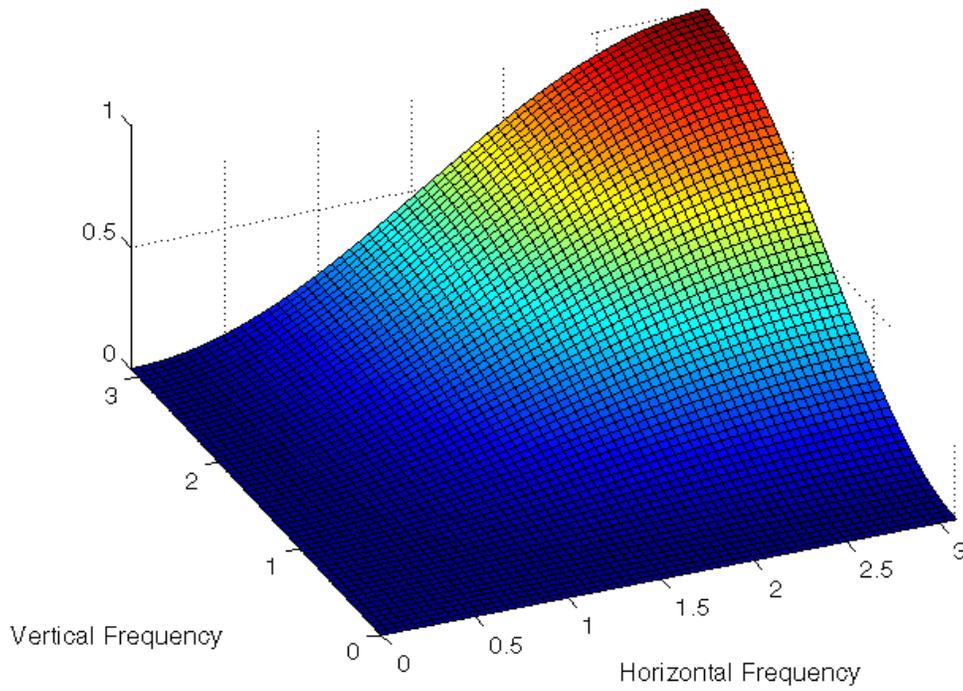


Figure 52 Frequency Response of $G_{11}(\omega_y)G_{11}(\omega_x)$ for Chui-Lian Multiwavelet

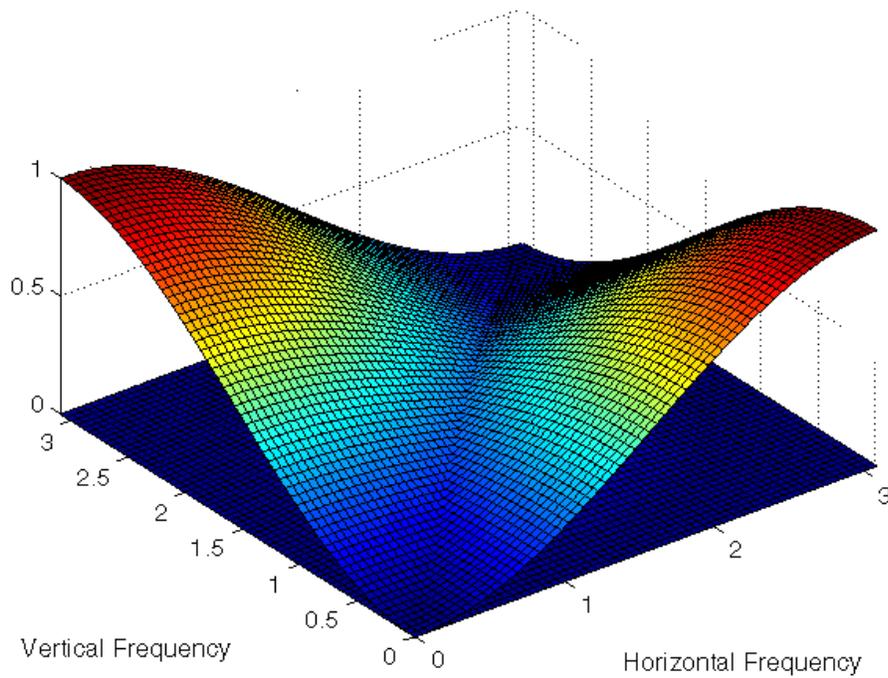


Figure 53 Frequency Responses of $G_{11}(\omega_y)H_{11}(\omega_x)$ and $H_{11}(\omega_y)G_{11}(\omega_x)$ for Chui-Lian Multiwavelet

Now, consider the question of the LLHH and HHH packets, designated in Figure 50. Again, when only the first basis function is considered, the ‘binary label’ method will suggest that the frequency ordering of the LLHH packet and HHHH packet would be reversed both horizontally and vertically, resulting in HHHH packet having a lower frequency coverage than LLHH packet alone both dimensions, as illustrated in Figures 54 and 55, respectively. This appears to be opposite to what we have analyzed based on the frequency coverage of the packet basis functions $\psi_1^{(k)}(\omega)$. The coupling between the two basis functions in each packet supplies high frequency content in these two packets such that no switching of frequency order is needed. We have examined this frequency ordering problem for other multiwavelets under consideration. The Cardinal Two-Balanced (Cardbal2) multiwavelet, the frequency spectra $(\psi_1^{(k)}(\omega), \psi_2^{(k)}(\omega))$, $(k=0, 1, 2, 3)$ are shown in Figure 56, which shows that the switching of frequency ordering of multiwavelet packet is needed, this is true for the multiwavelet GHM also, but not true for the SA4 multiwavelet, the frequency spectra $(\psi_1^{(k)}(\omega), \psi_2^{(k)}(\omega))$, $(k=0, 1, 2, 3)$ are shown in Figure 57.

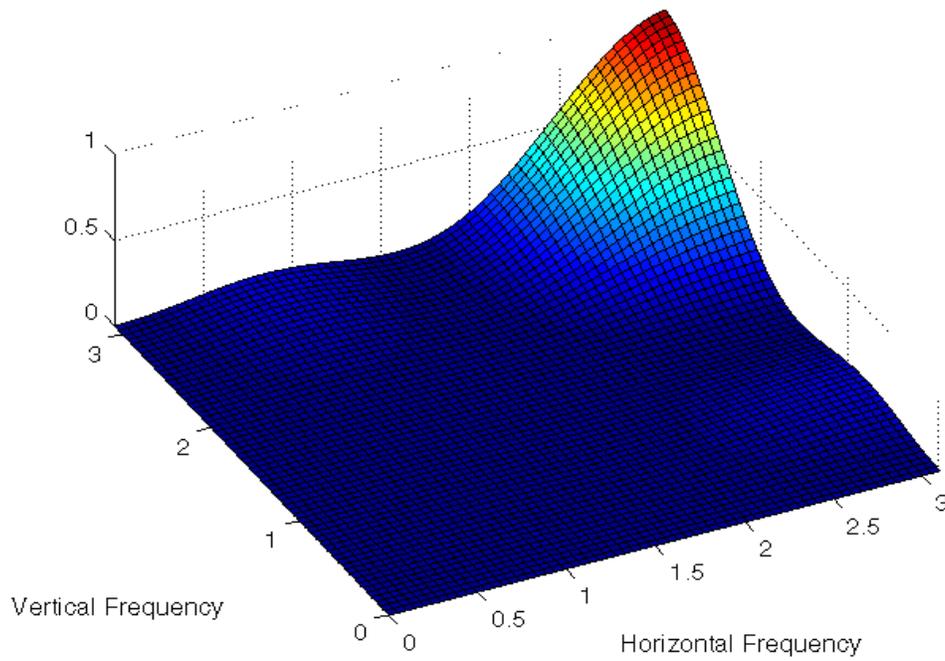


Figure 54 Frequency Response $H_{11}(\omega_y)H_{11}(2\omega_y)G_{11}(\omega_x)G_{11}(2\omega_x)$ of CL Multiwavelet Filters for LLHH Packet

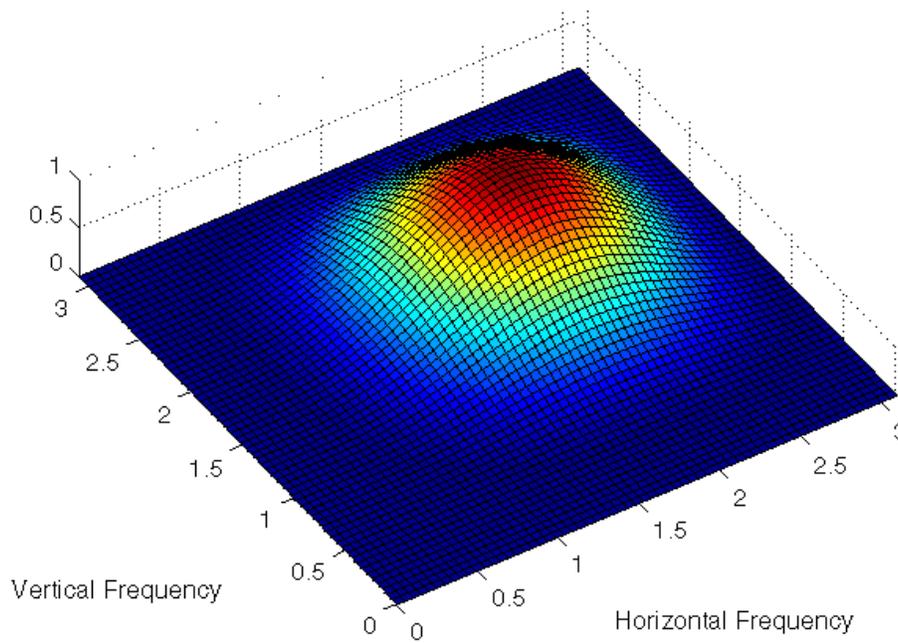


Figure 55 Frequency Response $G_{11}(\omega_y)G_{11}(2\omega_y)G_{11}(\omega_x)G_{11}(2\omega_x)$ of CL Multiwavelet Filters for HHHH Packet

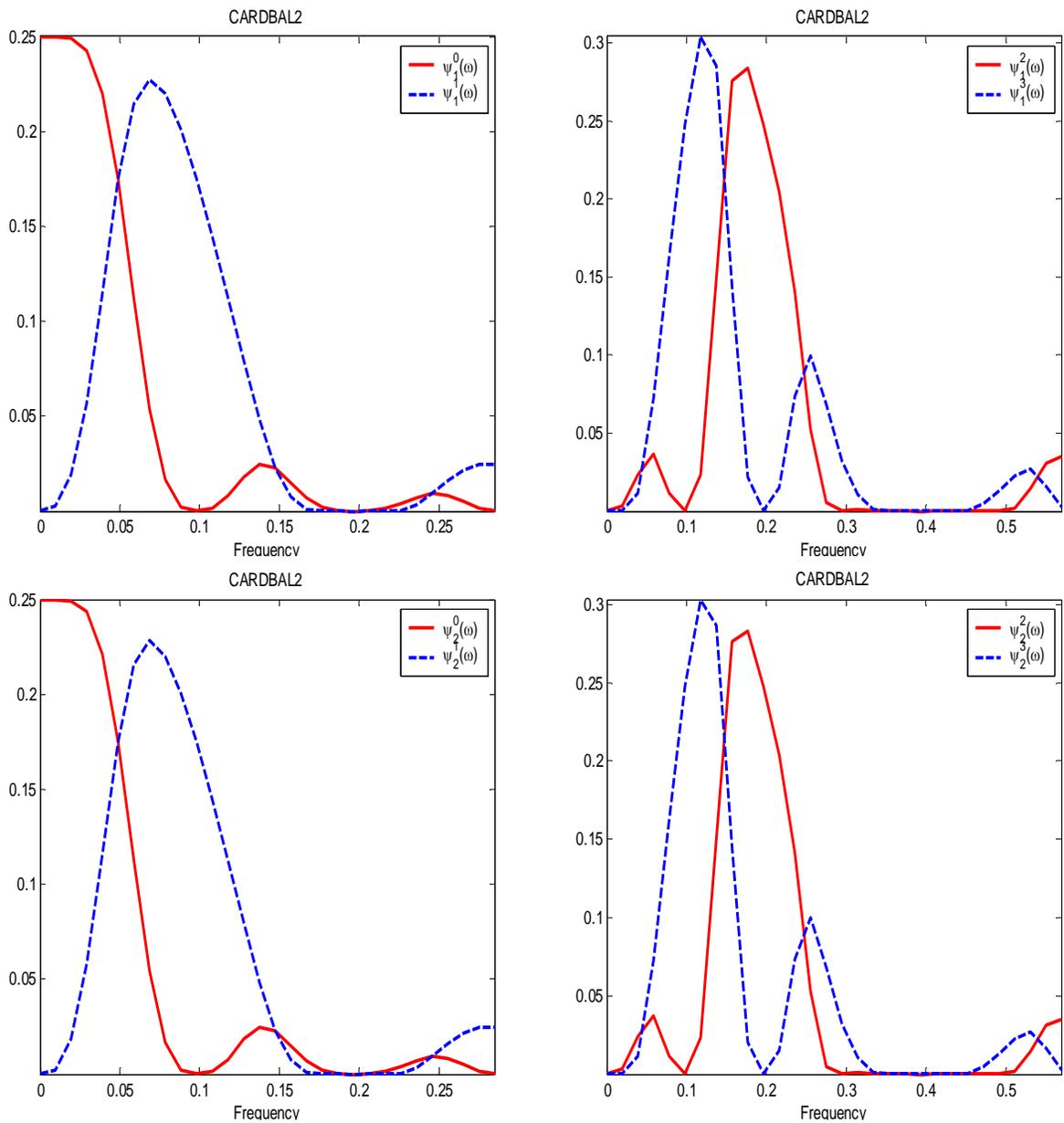


Figure 56 The Frequency Spectra of $\psi(k)(\omega)$ ($k=0,1,2,3$) of the Cardbal2 Multiwavelet

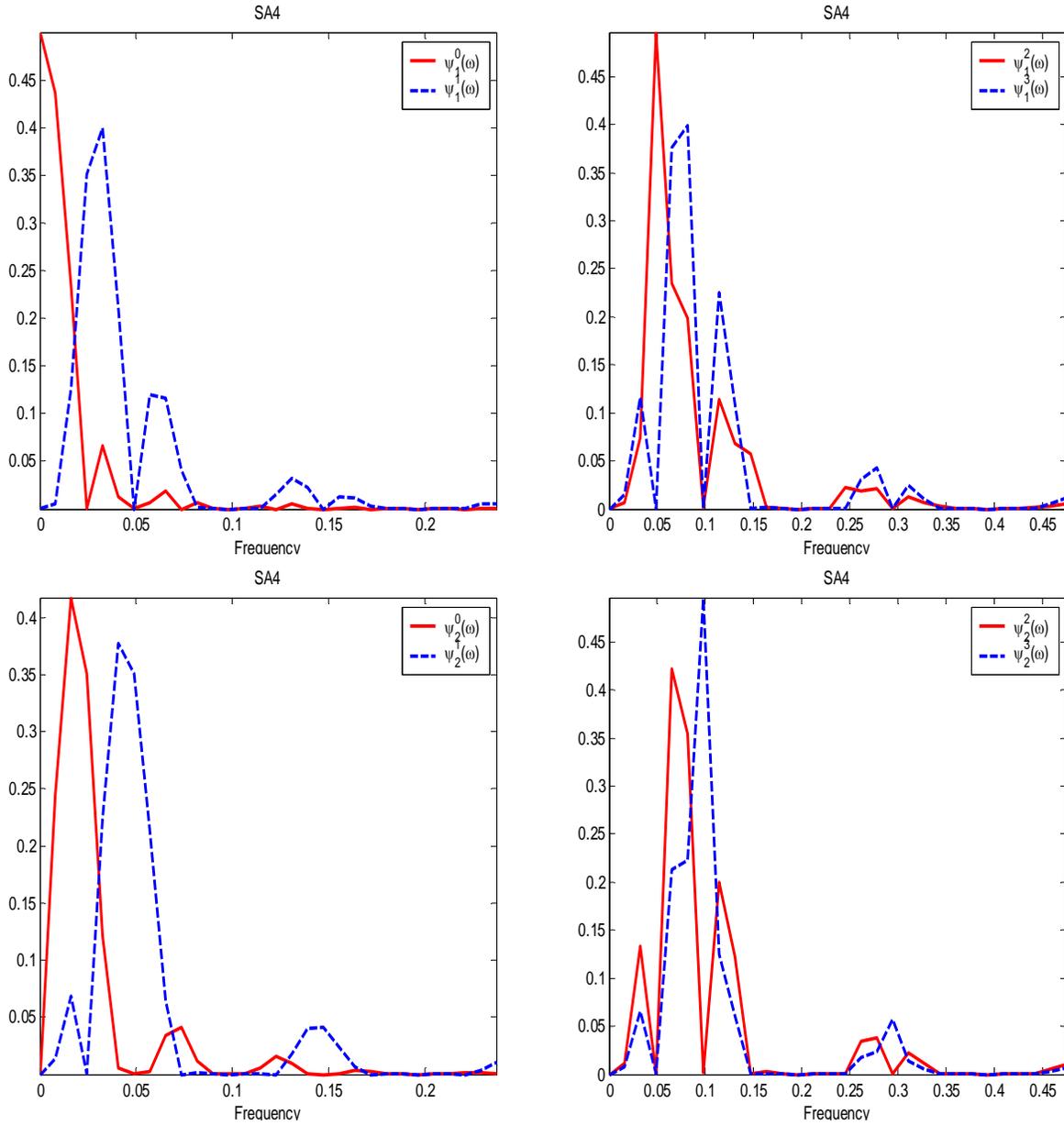


Figure 57 The Frequency Spectra of the $\psi(k)(\omega)$ ($k=0,1,2,3$) of the SA4 Multiwavelet

5.3 ORDER BY ENERGY

The second criteria of packet sorting by energy will allow the shuffling of packets between cousins or the modified parent-children relations as denoted in Figure 47. Since the SPIHT algorithm tests the significance of wavelet coefficients in each tree/set in the order of traversal,

so if we allow packets with higher energy to be switched with packets of lower energy, it is possible that the use of SPIHT in coding of wavelet packet coefficients can be made more efficient. We set the criterion for sorting the energy in each packet which is defined here as the summation of the absolute value of every coefficient, instead of its square, in that packet; i.e. the energy in a packet determines its distance from the root node. In order to minimize the number of comparisons in sorting, only the packets in the same modified parent-children subtree/set can be swapped, and the coarsest packet corresponding to $\psi^{(0)}(x)$ or LL in the 2-D case will not be tested since it generally contains most of the energy in a signal or image. There is an overhead cost associated with this type of sorting since the receiver can not reorganize the packets without any additional information from the server. To reduce this overhead, the server may first send a single bit to indicate whether there is any swapping of packets occurred, and then send the information on change of position of each packet. This criterion is also implemented in the experimental studies.

5.4 EXPERIMENT AND RESULTS

This experiment was designed to explore the effectiveness of multiwavelet packet on low bit-rate 3-D video compression and whether the shuffling of packets may contribute any performance advantage in coding coarse motion residuals. We compared the performance of multiwavelet packets, with and without shuffling; using the four medical endoscopy sequences (M08, M09, M10, and M11) used in the experiments reported in Chapter 4 with the same test parameters (two-level symmetric decomposition and a compression ratio of 1/40). Three multiwavelets were used, they were Chui-Lian multiwavelet (CL), the symmetric-asymmetric multiwavelet SA4, and cardinal balanced multiwavelet Cardbal2. In all the cases where multiwavelet packet

was used in the experiment, a complete-tree decomposition structure and its associated SPIHT structure which was presented earlier in this chapter were employed.

In frequency ordering portion of the experiment, we shuffled the packets according to their frequency order. For example, in the case of 2-D wavelet packet with two levels of decomposition, the packets LLHH and HHHH would be swapped during the encoding process because HHHH has a lower frequency order than LLHH. The result is that the distance of each packet from the root node would be determined by its frequency order and not by the decomposition order. The packets are re-shuffled back to their decomposition order during the decoding and reconstruction process. In the ordering based on packet energy, shuffling occurs only if the order of the energy of each packet is different from its decomposition order. Also the overhead cost of energy-based shuffling were ignored in this experiment. The following tables show the experimental results.

Table 4 Average PSNR of 3-D Multiwavelet Coarse Motion Residuals Coding with 3D SPIHT Structures (CL)

Sequence	Multiwavelet	Multiwavelet Packet without Sorting	Multiwavelet Packet with Sorting by Frequency Order	Multiwavelet Packet with Sorting by Energy Sum
M08	27.31	27.03	(N/A)	27.03
M09	27.92	27.66	(N/A)	27.66
M10	27.07	26.89	(N/A)	26.89
M11	27.15	27.04	(N/A)	27.04

Table 5 Totaltotal Residuals summed over 32 Endoscopy Video Framverage PSNR of 3-D Multiwavelet Coarse Motion Residuals Coding with 3D SPIHT Structures (SA4)

Sequence	Multiwavelet	Multiwavelet Packet without Sorting	Multiwavelet Packet with Sorting by Frequency Order	Multiwavelet Packet with Sorting by Energy Sum
M08	27.57	27.49	(N/A)	27.49
M09	27.76	27.72	(N/A)	27.72
M10	26.92	26.87	(N/A)	26.87
M11	27.69	27.62	(N/A)	27.62

Table 6 Average PSNR of 3-D Multiwavelet Coarse Motion Residuals Coding with 3D SPIHT Structures (Cardbal2)

Sequence	Multiwavelet	Multiwavelet Packet without Sorting	Multiwavelet Packet with Sorting by Frequency Order	Multiwavelet Packet with Sorting by Energy Sum
M08	26.93	27.21	27.21	27.21
M09	27.75	27.94	27.94	27.94
M10	26.91	27.09	27.09	27.09
M11	27.74	27.91	27.91	27.91

The results given in Tables 4,5 and 6 showed that the performances of Chui-Lian and SA4 3-D multiwavelet packets coding using SPIHT are slightly below that of the 3-D multiwavelet with SPIHT in these four sequences. However, the performance of the Carbal2 multiwavelet in wavelet packets coding actually improved on the 3-D multiwavelet by about 0.2dB on each sequence. A reasonable interpretation of the observation is that Cardbal2, having a longer filter length, has narrower frequency bands with smaller overlap between bands; this implies that significant coefficients are likely compactly distributed in packets at a certain

decomposition level thus requiring less bits to represent. However, such performance improvement in wavelet packet coding may be lost at a higher decomposition level where more packets may be involved in the spatial orientation trees. This interpretation is similar to the observation made by Martin and Bell that there exists an incompatibility between the SPIHT and wavelet packets coding in their study of multiwavelet packet on image compression [4].

The results given in the three tables also do not give support the desirability of shuffling packets. Any shuffling, if occurred at all, did not contribute to the overall performance at the current set of experimental parameters (2 levels of decomposition). It is yet to be seen whether the effects of shuffling packets may be observable if the number of levels of decomposition is increased.

In our experimental study of 3-D multiwavelet video coding, the level of decomposition was limited to 2 due the small number of frames included in each video segment for reducing the effects of frames containing zooming, panning and rotation. In the next experiment, we studied the use of 2-D multiwavelets and increased the number of decomposition levels to 3, 4 and 5 in the spatial domain. We used the coarse motion residual frame with the largest total energy from each of the four sequences used in the last experiment, and tested the effect of 2-D multiwavelet packet coding when more decomposition levels are used, but still at the compression ratio of 1/40. Tables 7,8, and 9 show the results of this experiment. Although these results of the 2-D multiwavelet coding of a single residual frame may not be directly translatable to the coding of a 3D block of frames, they may be taken as a suggestion. It is noted that multiwavelets with regular SPIHT obtained gains in PSNR as the number of level of decomposition increased, but the same was not true in the cases of multiwavelet packet coding. The performances of all three multiwavelet packets compared favorably with multiwavelet up to the decomposition level 4, but

a drop was observed at decomposition level 5 for all three multiwavelet packets. This performance drop fits the interpretation given earlier in the 3-D multiwavelet packets coding. It was also noted in the results that packet shuffling yielded neither gain nor loss at decomposition level 3. But at higher decomposition levels, it was observed that frequency sorting did not improve compression results at low bitrate of 0.2bpp regardless of the level of decomposition, but sorting by energy did improve slightly the results of multiwavelet packet, but these improvements did not take into account the overhead cost of having to send the sorted order of packets to the receiver.

Table 7 PSNR of 2-D Multiwavelet Coarse Motion Residuals Coding with 2D SPIHT (Chui-Lian) at 0.2bpp

Source Sequence of Frame	Method	Decomposition Level		
		3	4	5
M08	Multiwavelet	25.25	26.17	26.58
	Multiwavelet Packet	25.24	26.46	26.33
	Packet with Frequency Sort	(N/A)	(N/A)	(N/A)
	Packet with Energy Sort	25.25	26.47	26.35
M09	Multiwavelet	25.56	27.31	27.84
	Multiwavelet Packet	25.63	27.59	27.26
	Packet with Frequency Sort	(N/A)	(N/A)	(N/A)
	Packet with Energy Sort	25.63	27.61	27.25
M10	Multiwavelet	25.31	26.25	26.86
	Multiwavelet Packet	25.31	26.51	26.50
	Packet with Frequency Sort	(N/A)	(N/A)	(N/A)
	Packet with Energy Sort	25.31	26.52	26.51
M11	Multiwavelet	24.39	25.73	25.97
	Multiwavelet Packet	24.39	25.90	25.89
	Packet with Frequency Sort	(N/A)	(N/A)	(N/A)
	Packet with Energy Sort	24.39	26.00	25.89

Table 8 PSNR of 2-D Multiwavelet Coarse Motion Residuals Coding with 2D SPIHT (SA4) at 0.2bpp

Source Sequence of Frame	Method	Decomposition Level		
		3	4	5
M08	Multiwavelet	25.50	27.08	27.37
	Multiwavelet Packet	25.63	27.08	27.09
	Packet with Frequency Sort	(N/A)	(N/A)	(N/A)
	Packet with Energy Sort	25.63	27.12	27.12
M09	Multiwavelet	25.98	28.00	28.30
	Multiwavelet Packet	26.03	28.18	27.87
	Packet with Frequency Sort	(N/A)	(N/A)	(N/A)
	Packet with Energy Sort	26.03	28.20	27.87
M10	Multiwavelet	25.58	26.99	27.32
	Multiwavelet Packet	25.66	27.22	27.15
	Packet with Frequency Sort	(N/A)	(N/A)	(N/A)
	Packet with Energy Sort	25.66	27.22	27.18
M11	Multiwavelet	24.75	26.33	26.82
	Multiwavelet Packet	24.78	26.58	26.51
	Packet with Frequency Sort	(N/A)	(N/A)	(N/A)
	Packet with Energy Sort	24.78	26.60	26.49

Table 9 PSNR of 2-D MultiwaveletCoarse Motion Residuals Coding with 2D SPIHT (Cardbal2) at 0.2bpp

Source Sequence of Frame	Method	Decomposition Level		
		3	4	5
M08	Multiwavelet	25.30	26.83	26.91
	Multiwavelet Packet	25.53	27.13	27.06
	Packet with Frequency Sort	25.53	27.11	27.04
	Packet with Energy Sort	25.59	27.13	27.05
M09	Multiwavelet	26.28	27.70	27.91
	Multiwavelet Packet	26.65	28.15	28.06
	Packet with Frequency Sort	26.65	28.13	28.04
	Packet with Energy Sort	26.69	28.15	28.05
M10	Multiwavelet	25.46	26.79	27.02
	Multiwavelet Packet	25.64	27.18	27.30
	Packet with Frequency Sort	25.64	27.16	27.27
	Packet with Energy Sort	25.66	27.20	27.32
M11	Multiwavelet	25.20	25.99	26.53
	Multiwavelet Packet	25.26	26.67	26.65
	Packet with Frequency Sort	25.26	26.66	26.62
	Packet with Energy Sort	25.27	26.68	26.67

5.5 SUMMARY

We have examined the results of applying 3-D multiwavelet packets and using 3-D SPIHT on wavelet packets in video coding. The results of 3-D multiwavelet packets coding using multiwavelet Carbal2 gained an average improvement of 0.2dB over the regular wavelet coding, while CL and SA4 both suffered slight loss (0.2dB for CL, and 0.06dB for SA4). In the results of 2-D coarse-motion-residual compression using multiwavelet packets, all three multiwavelets packets gained some improvements over the regular multiwavelet coding at decomposition levels 3 and 4, especially Carbal2, but their performance started to deteriorate beyond decomposition level 4. Our interpretation of is that is due to narrower frequency bands of the Carbal2 filter so that the energy of significant coefficients are more compactly represented rather than disbursed over several packets. Our experiment in 2-D wavelet packet coding revealed that there is a limit to the improvement by Carbal2 when the decomposition level increased beyond 4. We found no significant contribution with multiwavelet packets subband shuffling in the results.

6.0 CONCLUSIONS AND FUTURE RESEARCH DIRECTIONS

In this dissertation research, we investigated the application of three-dimensional multiwavelet to color video compression which can simultaneously provide scalability and flexibility of quantization in both spatial and temporal domains. Two major contributions are summarized in the following paragraphs.

The first contribution is the development of a video coder using orthogonal 3-D multiwavelet and 3-D SPIHT for coding video and medical sequences. This approach has not been done before, and we have shown, by simulation, that 3-D compression using multiple scaling functions and multiple wavelets has better performance at low bit-rates than comparable single wavelet. We have also shown, by the same simulation, that performance 3D compression in the luminance domain may not be directly translated to the chrominance domain as behavior in some cases, as their rate of distortion in their respective entropies varies differently against the same quantizing step. The Geronimo-Hardin-Massopust (GHM) multiwavelet and the Chui-Lian (CL) multiwavelet, both having the approximation order of two, were used in the simulation in which video segments with 64 frames each were extracted from nine video sequences. It was found that at the low bit-rate of 0.1bpp, that is, compression ratio of 1/80, the 3-D multiwavelet CL outperformed the 3-D single wavelet DB4 by 1dB in the luminance domain and more than 4.5 dB in the chrominance domains; but that gain decreases as the bit-rate increases.

The second contribution is the development of the “coarse-motion-residual” 3-D multiwavelet video coder based on the coder discussed above by incorporating multiwavelet

transform of the errors of predicted motion trajectories as an unbiased platform to evaluate the performance of 3-D against 2-D in wavelet-based video compression. The performance measurements of 3-D vs. 2-D in the literatures are mostly based on codec specifically designed for each respective method. However, the “coarse-motion-residual” model was developed by incorporating 2-D coding methods into a 3-D framework, which allows an “apple-to-apple” comparison between 2-D and 3-D wavelet-based video coding. We utilize a key frame extracted from the video segment as the common reference for evaluating motion trajectory information of every other frame with respect to the key frame, and compute the change in intensity values (residuals). The residuals are then coded using the 3-D multiwavelet coder discussed above. A modified version of 3-D SPIHT was developed by taking advantage of the spatio-temporal relationship of the “coarse-motion-residuals”. A key frame selection procedure was also developed with the goal of reducing total residuals. Our experimental study based on multiwavelet CL revealed that the key frame selection procedure reduced total residuals, on the average, from 29.6×10^6 per frame to 16.8×10^6 per frame which is more comparable to the average of 13.8×10^6 obtained by 2-D wavelet-based coder using DB4. Two more orthogonal multiwavelets, Tham’s SA4 and Selesnick’s Cardbal2, were also used in the experiment. In spite of higher average residuals, the 3-D multiwavelet “coarse-motion-residuals” video coding outperformed the 2-D motion residuals coding in every case by 0.3dB to 4.88dB(with multiwavelet CL), 0.2dB to 4.69dB(with multiwavelet SA4), and 0.19dB to 5.12dB (with multiwavelet Cardbal2), even at the low bit-rate of 0.2bpp. The use of the modified SPIHT further enhanced the advantage of 3-D “coarse-motion-residuals” video coder. A lesser contribution from the simulations is that performances 3-D multiwavelets in compression can not be inferred from those of 2-D multiwavelets.

Lastly, we have explored the plausibility of using multiwavelet packets and subband shuffling in 3-D and 2-D in order to improve “coarse-motion-residuals” coding. By examining the frequency content of the $(\psi_1^{(k)}(x), \psi_2^{(k)}(x))$, ($k=0,1,2,3$) of each multiwavelet under consideration, we found that some multiwavelets do not require packet switching in the ordering of frequency; this observation was taken into consideration in the experimental studies. The performance improvement of 3-D multiwavelet packet over regular multiwavelet was observed only when balanced multiwavelet Cardbal2 was used, that has a longer filter length. We interpreted it as due to its narrower frequency bands thus keeping more energy compaction of significant coefficients. But this performance improvement is limited to a small number of decomposition levels.

6.1 SUGGESTIONS FOR FUTURE RESEARCH

The following problems are suggested for future research:

- 1 Optimize bit allocation across color domains: We have demonstrated in experiments that allocating the same number of bits per pixel for each color domain resulted in different PSNR values. This suggests an optimization problem for bit allocation across the color domains.
- 2 Develop an optimal coding strategy for multiwavelet packet in 3-D video compression: We have observed in the experiments that the SPIHT coder is not well-suited to multiwavelet packet coding, so a new scheme should be developed for wavelet packet-based compression and incorporate analysis functions as part of the scheme.
- 3 Building interactivity into 3-D wavelet coder: In a true 3D environment, users may wish to examine certain portions of video (or medical) sequences in detail. The 3D video coder can

be designed to efficiently send coefficients in regions determined by users. Furthermore, separate coding of video foreground and background can improve interactivity.

APPENDIX

RESULTS OF 0.1 BIT-PER-PIXEL VS. 0.25 BIT-PER-PIXEL 3D COMPRESSION

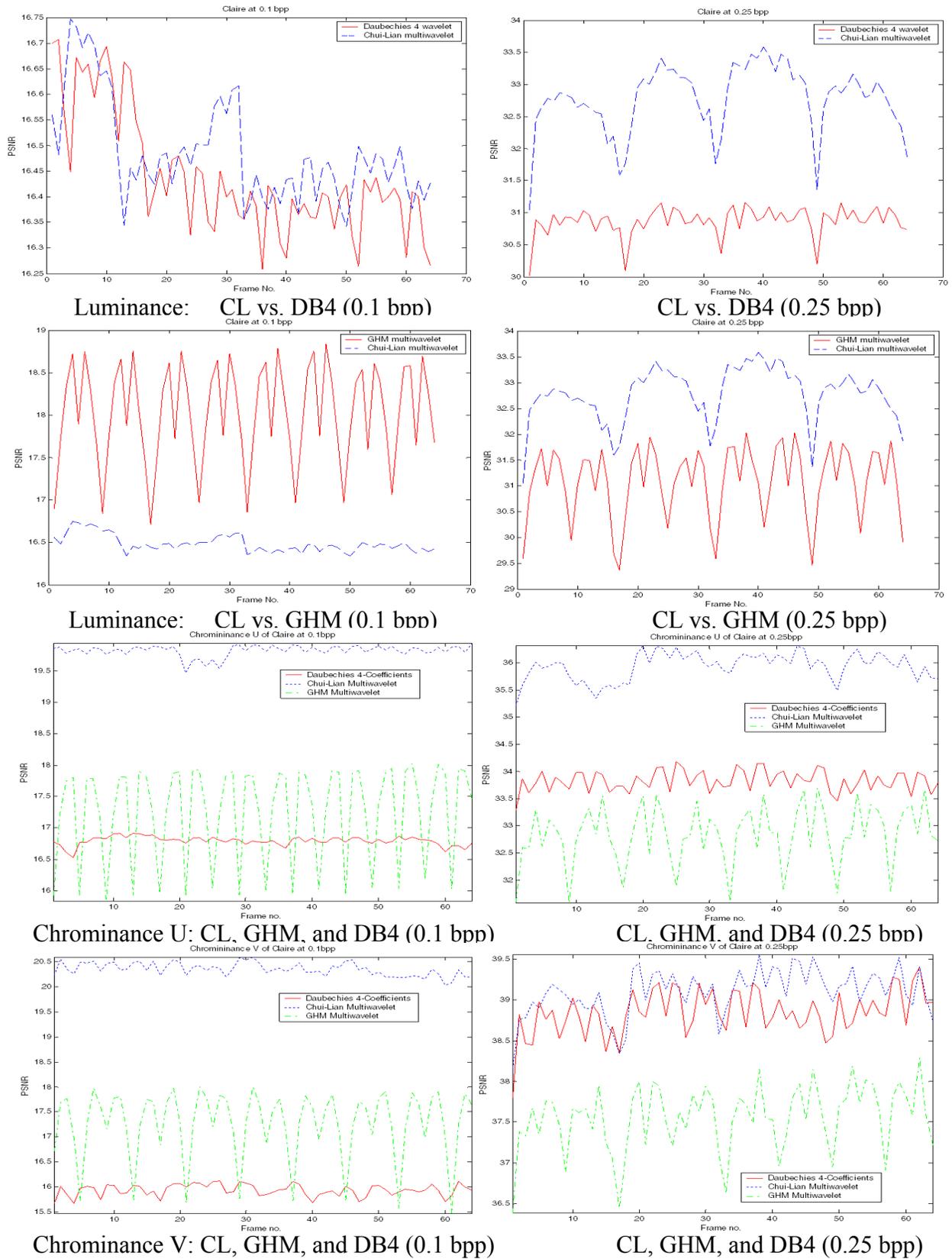


Figure 58 CL, GHM multiwavelets, and DB4 PSNR at 0.1 bpp and 0.25 bpp for Claire

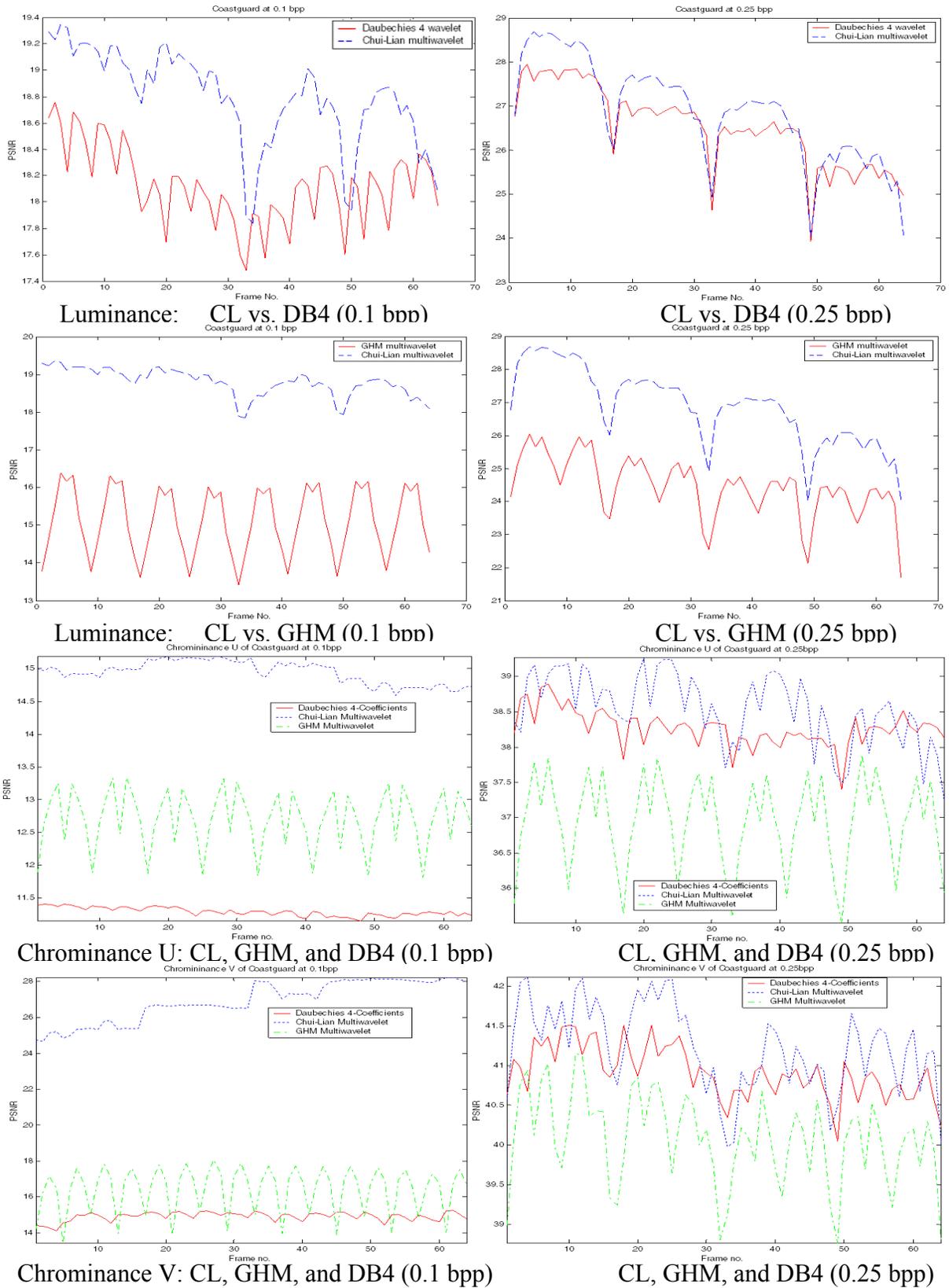


Figure 59 CL, GHM multiwavelets, and DB4 PSNR at 0.1 bpp and 0.25 bpp for Coastguard

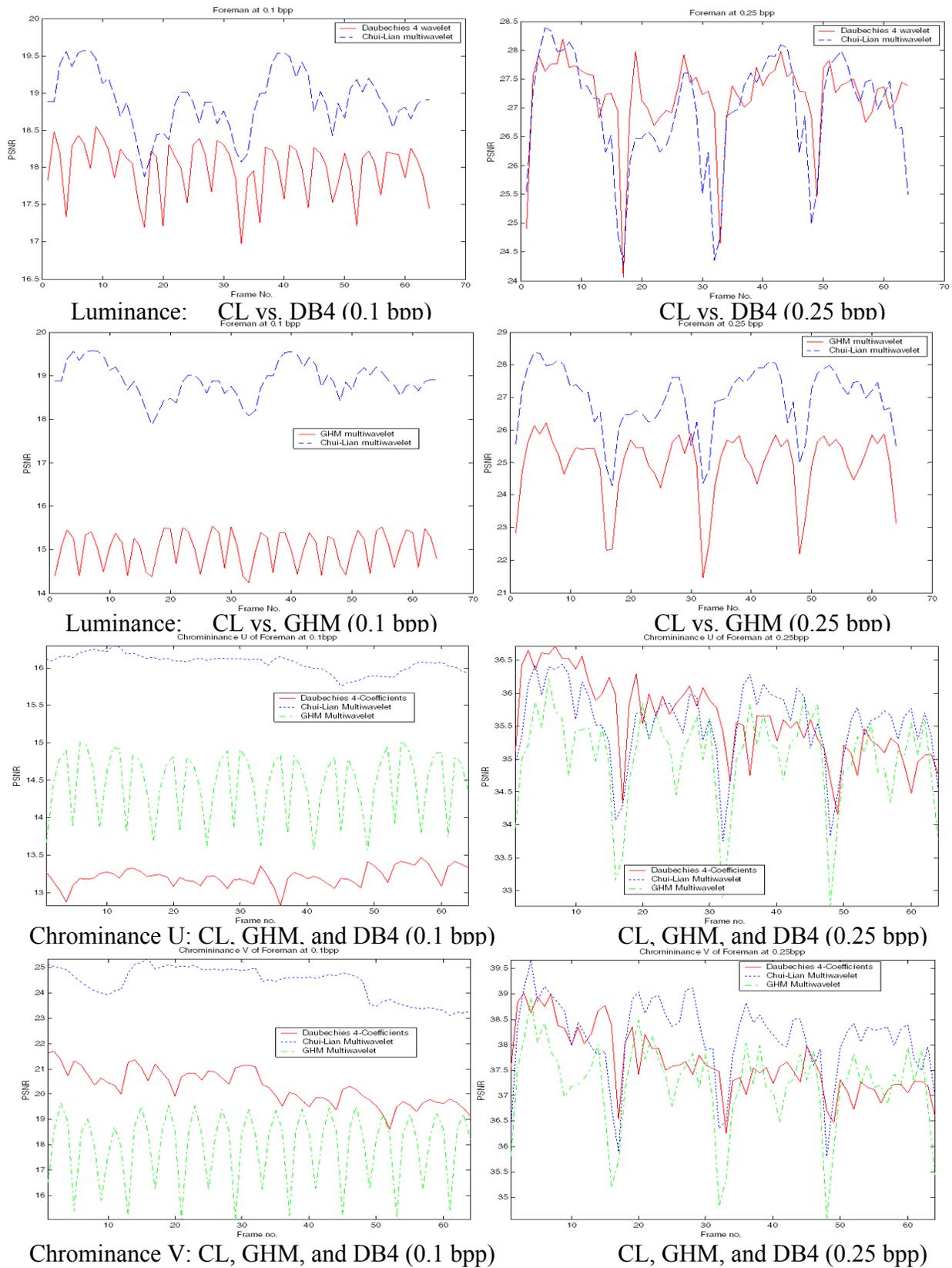


Figure 60 CL, GHM multiwavelets, and DB4 PSNR at 0.1 bpp and 0.25 bpp for Foreman

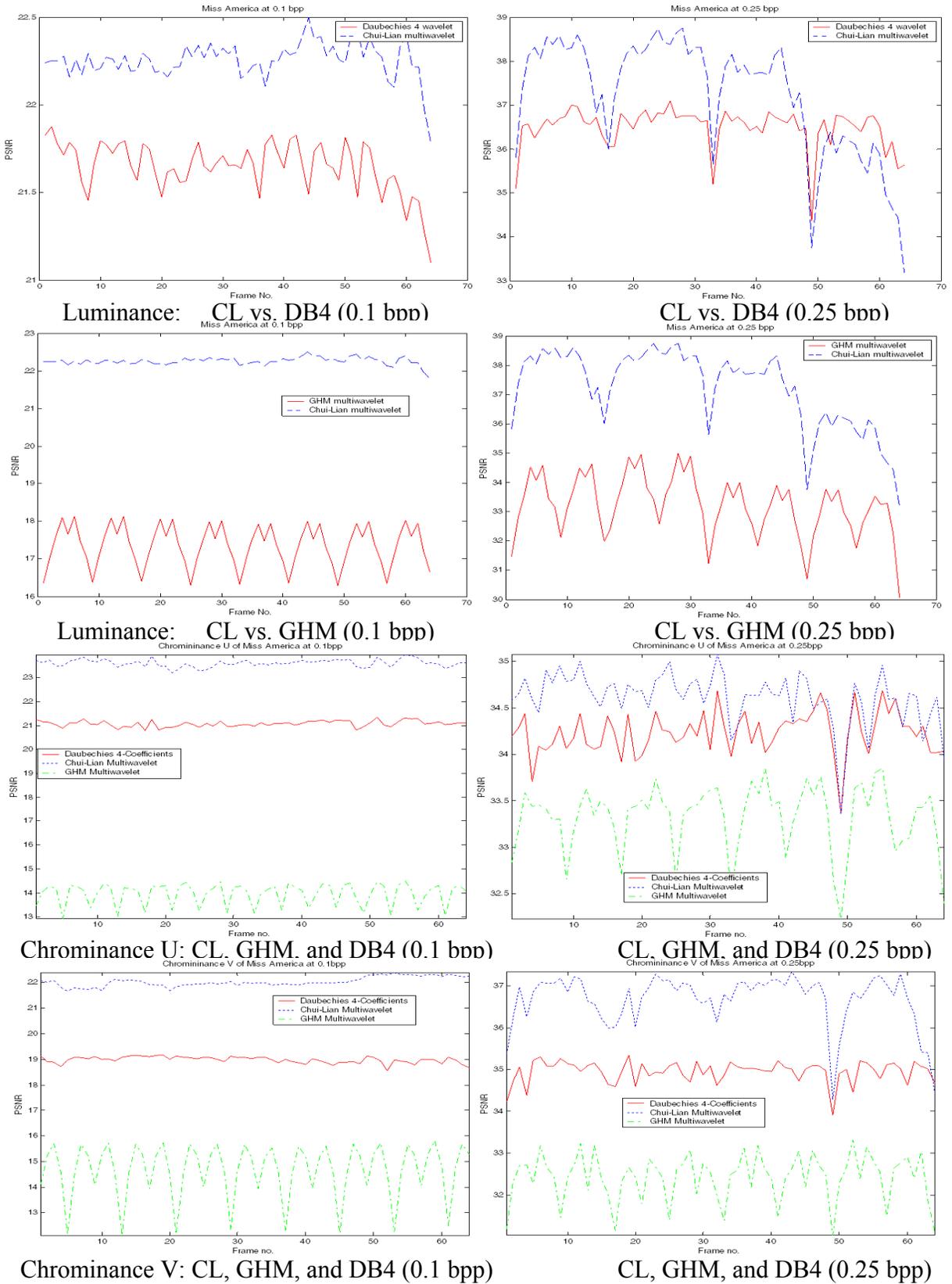


Figure 61 CL, GHM multiwavelets, and DB4 PSNR at 0.1 bpp and 0.25 bpp for Miss America

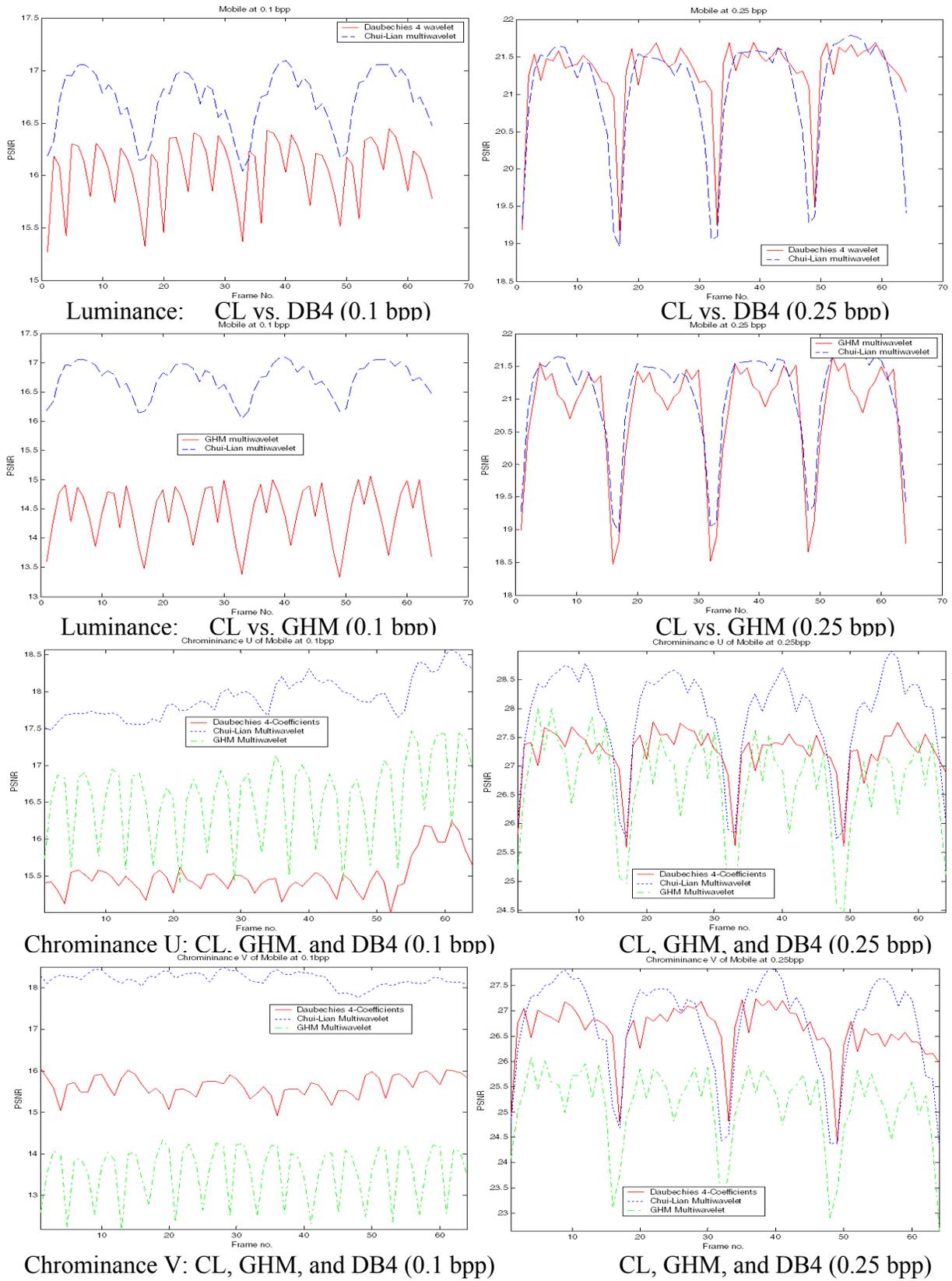


Figure 62 CL, GHM multiwavelets, and DB4 PSNR at 0.1 bpp and 0.25 bpp for Mobile

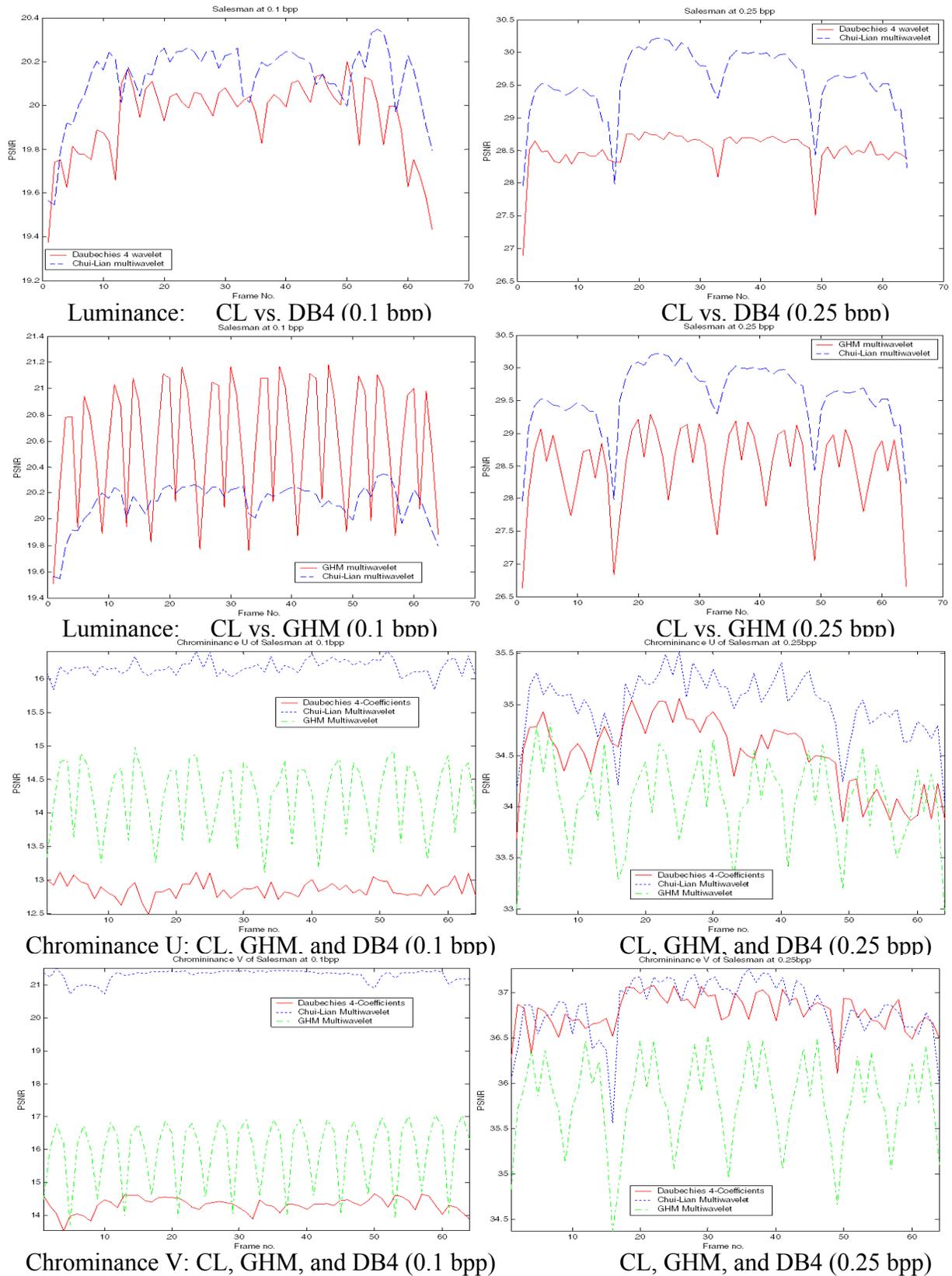


Figure 63 CL, GHM multiwavelets, and DB4 PSNR at 0.1 bpp and 0.25 bpp for Salesman

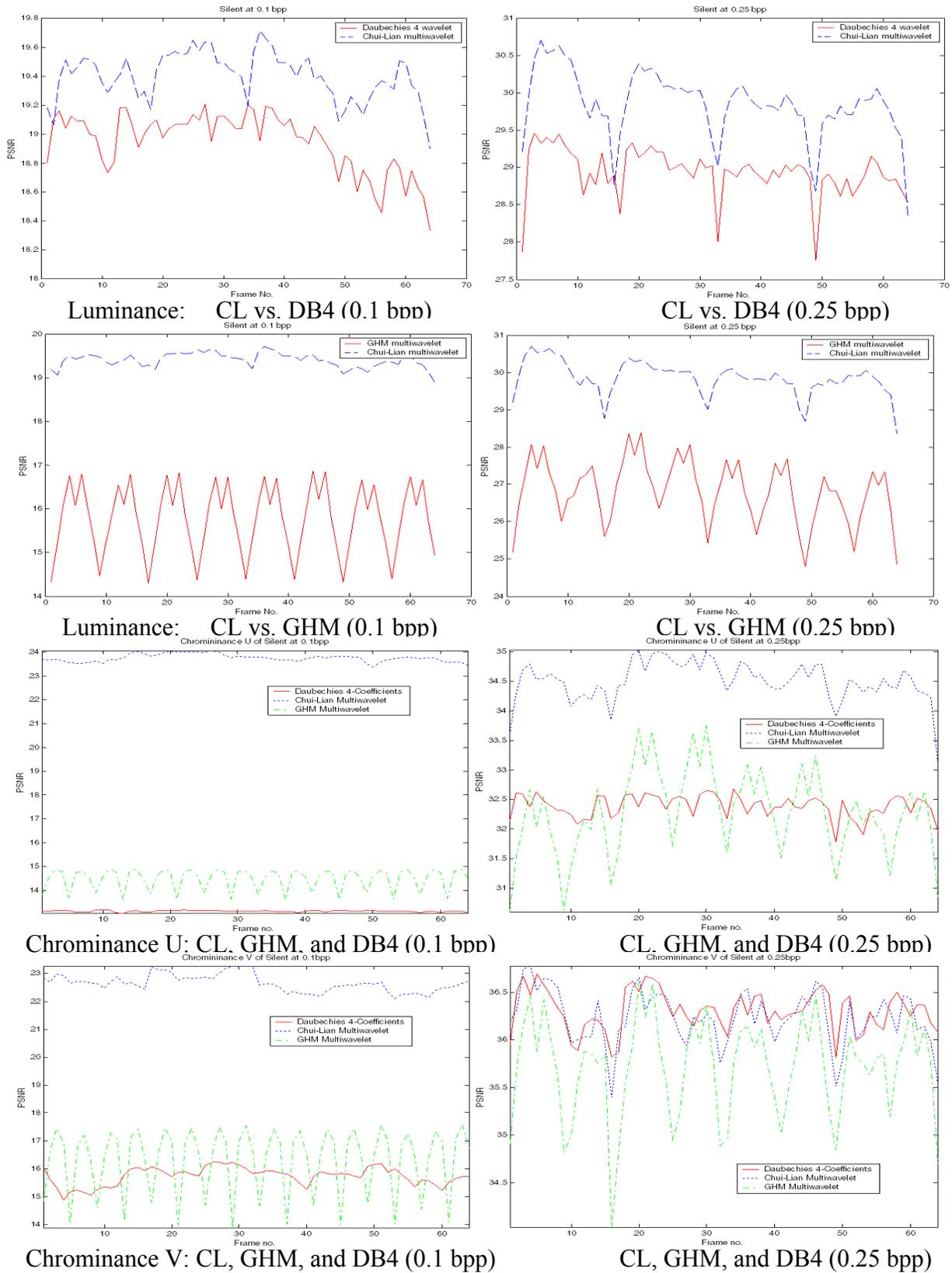


Figure 64 CL, GHM multiwavelets, and DB4 PSNR at 0.1 bpp and 0.25 bpp for Silent

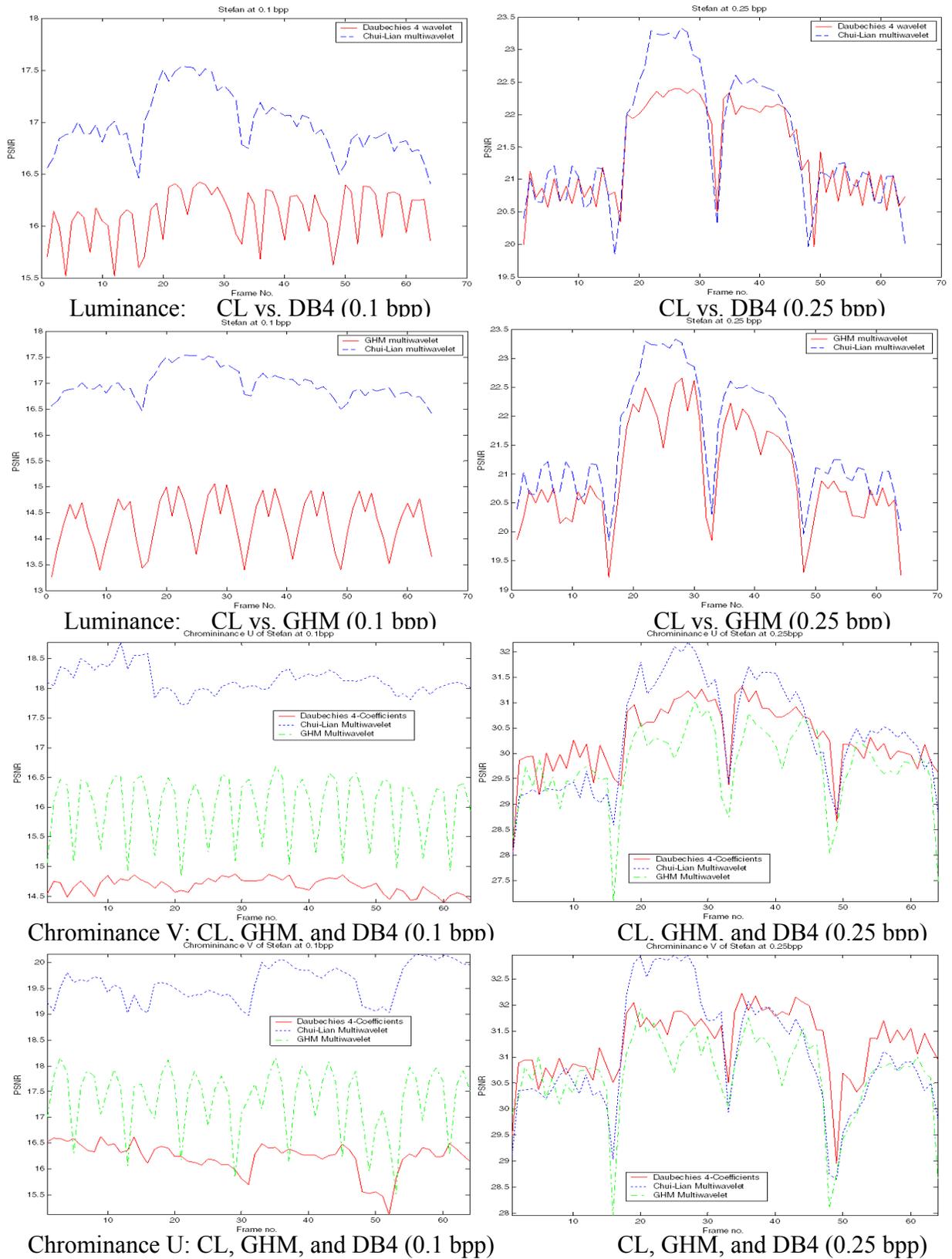


Figure 65 CL, GHM multiwavelets, and DB4 PSNR at 0.1 bpp and 0.25 bpp for Stefan

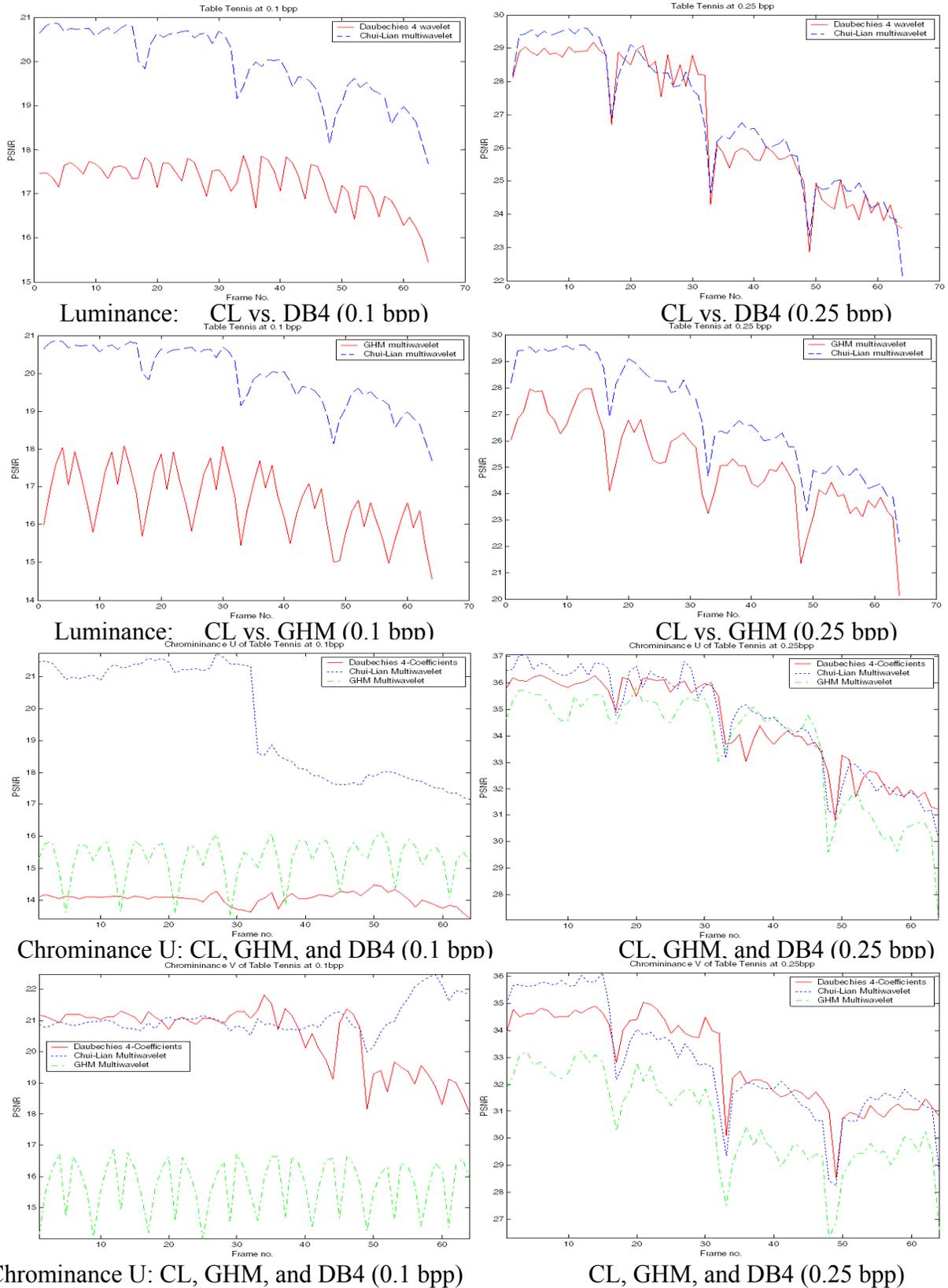


Figure 66 CL, GHM, and DB4 PSNR at 0.1 bpp and 0.25 bpp for Table Tennis

BIBLIOGRAPHY

1. S.M. Ciurel, V. Buzuloiu, and V. Patrascu, "Multiwavelets in MRI", Proceedings of International Symposium on Signals, Circuits and Systems, vol. 1, pp. 149-152. July, 2003.
2. S. Rout, S, and A. E. Bell, "Color image compression: multiwavelets vs. single wavelets", Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, vol. 4, pp. 3501-3504. May 2002
3. L.R. Iyer, and A.E. Bell, "Improving image compression performance with balanced multiwavelets", Proceedings of the Thirty-Fifth Asilomar Conference on Signals, Systems and Computers, vol. 1, pp. 773 -777. Nov. 2001
4. M. B. Martin and A.E. Bell, "New image compression techniques using multiwavelets and multiwavelet packets", IEEE Transactions on Image Processing, vol. 10, pp. 500 - 510. April 2001
5. M. Cotronei, D. Lazzaro, D, L.B. Montefusco, and L. Puccio, "Image compression through embedded multiwavelet transform coding", IEEE Transactions on Image Processing, vol. 9, pp. 184 -189. Feb. 2000
6. G. Lin, and Z-M Liu, "The application of multiwavelet transform to image coding", Proceedings of the sixth IEEE International Conference on Electronics, Circuits and Systems, vol. 1, pp. 445 -448. Sept. 1999
7. G. Lin, and Z-M Liu, "The application of multiwavelet transform to image coding", IEEE Transactions on Image Processing, vol. 9, pp. 270 -273. Feb. 2000
8. K-W Cheung, C-H Cheung, and L-M Po, "A novel multiwavelet-based integer transform for lossless image coding", Proceedings of International Conference on Image Processing, vol. 1, pp. 444 -447. 1999
9. V. Strela, P.N. Heller, G. Strang, P. Topiwala, and C. Heil, "Application of multiwavelet filterbanks to image processing", IEEE Transactions on Image Processing, vol. 8, pp. 548 -563. April 1999
10. J. Chen, J. Zhou, S. Yu, Q. Xiao, and J. Xu, "Application of symmetric orthogonal multiwavelets and prefilter technique for Image compression", Journal of Computer Science and Technology, vol. 3, pp. 47-53. April 2003

11. C.Y-F Ho, B. W-K Ling, T.P-L Wong, A.Y-P Chan, and P. K-S Tam, "Fuzzy multiwavelet denoising on ECG signal", *Electronics Letters*, vol. 39, pp. 1163 -1164, Aug. 2003
12. W. Ling, "Orthogonal multiwavelets transform for image denoising", *Proceedings of the International Conference on Signal Processing*, vol. 2, pp. 987 -991. Aug. 2000
13. E. Bala, and A. Ertuzun, "Applications of multiwavelet techniques to image denoising", *Proceedings of International Conference on Image Processing*, vol. 3, pp. 24-28, June 2002
14. T. D. Bui, and G. Chen, "Translation-invariant denoising using multiwavelets", *IEEE Transactions on Signal Processing*, vol. 46, pp. 3414 -3420. Dec. 1998
15. G.Y. Chen, and T.D. Bui, "Multiwavelets denoising using neighboring coefficients", *IEEE Signal Processing Letters*, vol. 10, pp. 211 -214. July 2003
16. X. Wang, "Nonlinear multiwavelet transform based soft thresholding", *Proceedings of the 2000 IEEE Asia-Pacific Conference on Circuits and Systems*, pp. 775 -778. Dec. 2000
17. S. Bacchelli, and S. Papi, "Matrix Thresholding for Multiwavelet Image Denoising", *Numeric Algorithms*, vol. 23, pp. 41-52. Aug. 2003
18. K. Kazemi and H.A. Moghaddam, "Fusion of multifocus images using discrete multiwavelet ", *Proceedings of IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems*, pp. 167-172. July, 2003
19. T. Palfner, and E. Muller, "Effects of symmetric periodic extension for multiwavelet filter banks on image coding", *Proceedings of International Conference on Image Processing*, vol. 1, pp. 628 -632. 1999
20. K. Jafari-Khouzani, and H. Soltanian-Zadeh, "Multiwavelet grading of pathological images of prostate", *IEEE Transactions on Biomedical Engineering*, vol. 50, pp. 697-704. June 2003
21. G.W. Pan, K. Wang and B.K. Gilbert, "On multiwavelet-based finite-element method", *IEEE Transactions on Microwave Theory and Techniques*, vol. 51, pp. 148 -155. Jan. 2003
22. J-W Wang, "Multiwavelet packet transforms with application to texture segmentation", *Electronics Letters*, vol. 38, pp. 1021 -1023. Aug. 2002
23. C.A. Micchelli, "Reconstruction and Decomposition Algorithms for Biorthogonal Multiwavelets", *Multidimensional Systems and Signal Processing*, vol. 8, pp. 31-69. Jan. 1997

24. J.T. Miller, "Adaptive Multiwavelet Initialization", PhD Thesis, Univ. of Pittsburgh, Pittsburgh, PA. 1996
25. W. Kim, "Preconditioning Multiwavelet Systems for Image Compression", PhD Thesis, Univ. of Pittsburgh, Pittsburgh, PA. 1999.
26. W. Kim, and C.C. Li, "On Preconditioning Multiwavelet Systems for Image Compression", International Journal of Wavelets, Multiresolution and Information Processing, vol. 1, pp. 51-74. 2003
27. K. Attakitmongcol, D.P. Hardin, and D.M. Wilkes, "Optimal prefilters for the multiwavelet filter banks", Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, vol. 6, pp. 3613 -3616, May 2001
28. K. Attakitmongcol, D.P. Hardin, and D.M. Wilkes, "Multiwavelet prefilters. II. Optimal orthogonal prefilters", IEEE Transactions on Image Processing, vol. 10, pp. 1476 -1487. Oct. 2001
29. X. Yang, and L.C. Jiao, "Adaptive multiwavelet prefilter", Electronics Letters, vol. 35, pp. 11-13. Jan. 1999
30. T-C Hsung, M-C Sun, D.P-K Lun, and W-C Siu, "Symmetric prefilters for multiwavelets", Proceedings of IEE Vision, Image and Signal Processing, vol. 150, pp. 59 -68. Feb. 2003
31. J. C. Lagarias, and Y. Wang, "Orthogonality Criteria for Compactly Supported Refineable Functions and Refineable Function Vectors", The Journal of Fourier Analysis and Applications, vol. 6, pp. 153-170. 2000
32. J.S. Geronimo, D.P. Hardin, and P.R. Massopust, "Fractal function and wavelet expansions based on several scaling functions", Journal of Approximation Theory, vol. 78, pp. 373-401. 1994
33. C.K. Chui, and J.A. Lian, "A study of orthonormal multi-wavelets", Applied Numerical Mathematics, vol. 20, pp. 273-298. 1996
34. X. Li, "New results of phase shifting in the wavelet space", IEEE Signal Processing Letters, vol. 10, pp. 193-195. July 2003
35. H.Y.H. Chuang, D.P. Birch, L.C. Liu, J.C. Chien, S.P. Levitan and C.C. Li, "A High Speed Shift-Invariant Wavelet Transform Chip for Video Compression", Proceedings of IEEE Computer Society Annual Symposium on VLSI, pp. 125-131. April 2002
36. Y. Hui, C.W. Kok, T.Q. Nguyen, "Theory and design of shift-invariant filter banks and wavelets", Proceedings of Time-Frequency and Time-Scale Analysis, pp. 49 -52. June 1996

37. Y. Hui, C.W. Kok, T.Q. Nguyen, "Image compression using shift-invariant dyadic wavelets", Proceedings of International Conference on Image Processing, vol. 1, pp. 61 - 64. Oct. 1997
38. C. He, J. Dong, Y.F. Zheng, and Z. Gao, "Optimal 3-D coefficient tree structure for 3-D wavelet video coding", IEEE Transactions on Circuits and Systems for Video Technology, vol. 13, pp. 961-972. October 2003
39. C. Parisot, M. Antonini, and M. Barlaud, "3D scan-based wavelet transform and quality control for video coding", EURASIP Journal on Applied signal Processing, vol. 1, pp. 56-65. 2003
40. H. Man; R. L. de Queiroz, and M.J.T. Smith, "Three-dimensional subband coding techniques for wireless video communications", IEEE Transactions on Circuits and Systems for Video Technology, vol. 12, pp. 386-397. June 2002
41. M. Weeks, and M.A. Bayoumi, "Three-dimensional discrete wavelet transform architectures", IEEE Transactions on Signal Processing, vol. 15, pp. 2050-2063. Aug. 2002
42. M. Hashimoto, K. Matsuo, A. Koike, and Y. Nakajima, "Hierarchical structure for data transmission of volumetric medical images using three-dimensional wavelet transform", Nuclear Science Symposium Conference Record, vol. 3, pp. 1399 -1403. Nov. 2001
43. E. Moyano, F.J. Quiles, A. Garrido, T. Orozco-Barbosa, and J. Duato, "Efficient 3D wavelet transform decomposition for video compression", Proceedings of International Workshop on Digital and Computational Video, pp. 118-125. Feb. 2001
44. I.K. Levy, and R. Wilson, "Three dimensional wavelet transform video compression", IEEE International Conference on Multimedia Computing and Systems, vol. 2, pp. 924 - 928, June 1999
45. I.K. Levy, and R. Wilson, "Three-dimensional wavelet transform video coding using symmetric codebook vector quantization", IEEE Transactions on Image Processing, vol. 10, pp. 470 -475. March 2001
46. C. Zhang, H.X. Zheng, and D.B. Ge, "3-D wavelet-based multiresolution analysis", Proceedings of Fifth International Symposium on Antennas, Propagation and EM Theory, pp. 301 -304. Aug. 2000
47. N.V. Boulgouris, A. Leontaris, and M.G. Strintzis, "Wavelet compression of 3D medical images using conditional arithmetic coding", Proceedings of the 2000 IEEE International Symposium on Circuits and Systems, vol. 4, pp. 557-560. May 2000
48. Wang, Z. Xiong, P.A. Chou, and S. Mehrotra, "Three-dimensional wavelet coding of video with global motion compensation" Proceedings of Data Compression Conference, pp. 404-413. March 1999

49. J. Vass, and X. Zhuang, "Efficient three-dimensional wavelet codecs for networked video communications", Proceedings of International Conference on Image Processing, vol. 3, pp. 565-569. Oct. 1999
50. P. Campisi, M. Gentile, and A. Neri, "Three dimensional wavelet based approach for a scalable video conference system", Proceedings of International Conference on Image Processing, vol. 3, pp. 802 -806. Oct. 1999
51. S-J Choi, and J.W. Woods, "Motion-compensated 3-D subband coding of video", IEEE Transactions on Image Processing, vol. 8, pp. 155 -167. Feb. 1999
52. G.H. Lee, J.H. Song, and R-H Park, "Three-dimensional DCT/WT compression using motion vector segmentation for low bit-rate video coding", Proceedings of International Conference on Image Processing, vol. 3, pp. 456 -459, Oct. 1997
53. Z. Xiong, X. Wu, S. Cheng, and J. Hua, "Lossy-to-lossless compression of medical volumetric data using three-dimensional integer wavelet transforms", IEEE Transactions on Medical Imaging, vol. 22, pp. 459-470. March 2003
54. G. Menegaz, and J-P Thiran, "Three-dimensional encoding/two-dimensional decoding of medical data", IEEE Transactions on Medical Imaging, vol. 22, pp. 424 -440. March 2003
55. L. Luo, Y. Wu, J. Li, and Y-Q Zhang, "3-D wavelet compression and progressive inverse wavelet synthesis rendering of concentric mosaic", IEEE Transactions on Image Processing, vol. 11, pp. 802-816. July 2002
56. M. Weeks, and M.A. Bayoumi, "Three-dimensional discrete wavelet transform architectures", IEEE Transactions on Signal Processing, vol. 50, pp. 2050-2063. Aug. 2002
57. J. Zhu, "Image compression using wavelets and JPEG2000: a tutorial", Electronics & Communication Engineering Journal , vol. 14, pp. 112 -121. June 2002
58. X. Hang, N.L. Greenberg, T. Shiota, M.S. Firstenberg, and J.D. Thomas, "Compression of real time volumetric echocardiographic data using modified SPIHT based on the three-dimensional wavelet packet transform", Computers in Cardiology, pp. 123 -126. Sept. 2000
59. P. Luigi Dragotti, G. Poggi, A.R.P. Ragozini, "Compression of multispectral images by three-dimensional SPIHT algorithm", IEEE Transactions on Geoscience and Remote Sensing, vol. 38, pp. 416 -428. Jan. 2000
60. Said, and W.A. Pearlman, "A New Fast and Efficient image codec based on set partitioning in hierarchical trees", IEEE Transactions on Circuits and Systems for Video Technology, vol. 6, pp. 243-250. June 1996

61. B-J Kim, and W.A. Pearlman, "An embedded wavelet video coder using three-dimensional set partitioning in hierarchical trees (SPIHT)" Proceedings of Data Compression Conference, pp. 251-260. March 1997
62. B-J Kim, Z. Xiong, and W.A. Pearlman, "Low bit-rate scalable video coding with 3-D set partitioning in hierarchical trees (3-D SPIHT)", IEEE Transactions on Circuits and Systems for Video Technology, vol. 10, pp. 1374-1387. Dec. 2000
63. S. Cho, and W.A. Pearlman, "Improved error resilient embedded video coding", Proceedings of Data Compression Conference, pp. 83-92. March 2003
64. D. Mukherjee, and S.K. Mitra, "Vector SPIHT for embedded wavelet video and image coding", IEEE Transactions on Circuits and Systems for Video Technology, vol. 13, pp. 231-246. March 2003
65. A.A. Kassim, and W.S. Lee, "Embedded color image coding using SPIHT with partially linked spatial orientation trees", IEEE Transactions on Circuits and Systems for Video Technology, vol. 13, pp. 203-206. Feb. 2003
66. E. Khan, and M. Ghanbari, "Efficient SPIHT-based embedded colour image coding techniques", Electronics Letters, vol. 37, pp. 951-952. Jul 2001
67. A. Smolic and P. Kauff, "Interactive 3-D Video Representation and Coding Technologies", Proceedings of the IEEE Vol. 93, pp98-110, Jan. 2005
68. J-R Ohm, "Advances in Scalable Video Coding" Proceedings of the IEEE, vol. 93, pp. 42-56, Jan. 2005
69. J-C Chien and C.C. Li, "Skin color detection in low bit-rate 3-D multiwavelet-based videos", Proceedings of 17th International Conference on Pattern Recognition, vol 1, pp. 720-723, August 2004
70. S Zhao, Z. Xiong, X. Wang, "Optical Resource Allocation for Wireless Video over CDMA Network", IEEE Transactions on Mobile Computing, vol. 4, pp. 56-67, Jan.-Feb. 2005
71. J. Xu, Z. Xiong, S. Li and Y-Q Zhang, "Three-Dimensional Embedded Subband Coding with Optimized Truncation (3-D ESCOT)", Applied and Computational Harmonic Analysis, vol. 10, pp. 290-315, May 2001
72. L. Luo, F. Wu, S. Li, Z. Xiong, and Z. Zhuang, "Advanced Motion Threading for 3D Wavelet Video Coding", Signal Processing, vol. 19, pp. 601-616, Aug. 2004
73. T. Kronander, "New Results on 3-Dimensional Motion Compensated Subband Coding", Proceedings of the Picture Coding Symposium, pp. 8.5-10, 1990

74. L. Shen, H.H. Tan, and J.Y. Tham, "Some Properties of Symmetric-Antisymmetric Orthonormal Multiwavelets", IEEE Transactions on Signal Processing, Vol. 48, pp. 2161-2163, July 2000
75. J.Y. Tham, "Multiwavelets and Scalable Video Compression", PhD Dissertation, National University of Singapore, 2002.
76. I.W. Selesnick, "Balanced GHM-like Multiscaling Functions", IEEE Signal Processing Letters, vol. 6, pp. 111-112, May 1999
77. I.W. Selesnick, "Interpolating Multiwavelet Bases and the Sampling Theorem", IEEE Transactions on Signal Processing, vol. 47, pp. 1615-1621, June, 1999
78. I.W. Selesnick, "Balanced Multiwavelet Bases Based on Symmetric FIR Filters", IEEE Transactions on Signal Process, vol. 48, pp. 184-191, Jan. 2000
79. F. Keinert, "Wavelets and Multiwavelets", Chapman & Hall/CRC Publishing 2004