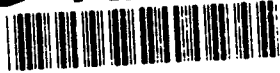


AD-A270 621



①

AFIT/DS/ENG/93D-12

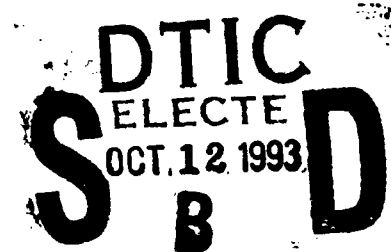
**A NON-HOMOGENEOUS, SPATIO-TEMPORAL,
WAVELET MULTIREOLUTION ANALYSIS
AND ITS APPLICATION TO THE ANALYSIS OF MOTION**

DISSERTATION

**Presented to the Faculty of the School of Engineering
of the Air Force Institute of Technology
Air University
In Partial Fulfillment of the
Requirements for the Degree of
Doctor of Philosophy**

**Thomas Joseph Burns, B.S.E.E., M.S.E.E.
Captain, USAF**

December, 1993



Approved for public release; distribution unlimited

142

93-23988



A NON-HOMOGENEOUS, SPATIO-TEMPORAL,
WAVELET MULTIREOLUTION ANALYSIS
AND ITS APPLICATION TO THE ANALYSIS OF MOTION

Thomas Joseph Burns, B.S.E.E., M.S.E.E.

Captain, USAF

Approved:

STEBS 10 Sept 93

Dennis W. Runk 10 SEP 93

Mark E. Opley 10 Sept 93

Michael C. Roggeman 10 Sep 93

Robert A. Calico, Jr

Robert A. Calico, Jr

Dean

DTIC QUALITY INSPECTED 2

Accession For	
NTIS GRA&I	☒
DTIC TAB	☐
Unannounced	☐
Justification	
By	
Distribution/	
Availability Codes	
Dist	Avail and/or Special
A-1	

Acknowledgements

It wasn't long after entering the PhD program that I discovered hard work comprises at least 90 percent of the doctoral research effort, with the remaining 10 percent split between inspiration, insight, creativity and luck. It comes as no surprise, then, that a doctoral student's most valuable resources are the colleagues, friends and family that unselfishly give of themselves to help the student lighten the workload. To these people, I am eternally grateful.

To my advisor and friend, Dr Steve Rogers, whose unflagging optimism sustained me through the tough times and whose razor-sharp insight provided the intellectual rudder that kept me on course. To my committee members, Dr Dennis Ruck and Dr Mark Oxley, who separately spent many hours patiently explaining and re-explaining fundamental concepts associated with the research and its implementation. To my Dean's Representative, Dr Mike Roggeman, who went above and beyond the call of duty. To my fellow doctoral student Ken Fielding, who served both as sounding board and sanity check for many of my ideas, and who was an invaluable research partner through all phases of the research. To my sponsor, Greg Power, for his steadfast support and encouragement. And to Dan Zambon, whose sense of humor and common sense injected a much needed dose of reality into an otherwise abstract and theoretical process.

Finally, and most importantly, I thank my wife, Anne, my children, Jacob and Joseph, and my mother and father. Their love, sacrifice and support made this goal a reality. All I've attained I owe to them and to my God.

Thomas Joseph Burns

Table of Contents

	Page
Acknowledgements	iii
List of Figures	vii
Abstract	xii
 I. Introduction	 1
1.1 Historical Background	1
1.2 Problem Statement and Scope	3
1.3 Dissertation Organization	6
 II. Background Material	 8
2.1 Introduction	8
2.2 Signal Analysis with a Wavelet Transform	8
2.2.1 The Continuous Wavelet Transform.	9
2.2.2 Signal Approximation With the Wavelet Series.	12
2.2.3 The Wavelet Multiresolution Analysis in $\mathbb{R}^1$ and $\mathbb{R}^2$	14
2.3 Traditional Methods for Computing Optical Flow	20
2.3.1 Feature Correspondence.	23
2.3.2 Spatio-Temporal Differentiation.	24
2.3.3 Fourier Phase Approach.	27
2.4 Motion Analysis Using a Spatio-Temporal Frequency Approach	28
2.5 Conclusion	34
 III. A Wavelet Multiresolution Analysis for $L_2(\mathbb{R}^3)$	 36
3.1 Introduction	36
3.2 Orthonormal Wavelet Basis	36

	Page
3.3 Discrete Multiresolution Decomposition Algorithm	40
3.4 Spatio-Temporal Filter Bank Representation	44
3.5 A Simple Application	47
3.6 Discussion	50
3.7 Conclusion	53
IV. A Non-Homogeneous, Motion-Oriented $L_2(\mathbb{R}^3)$ Wavelet Multiresolution Analysis . .	54
4.1 Introduction.	54
4.2 A Non-Homogeneous Wavelet Multiresolution Analysis for $L_2(\mathbb{R}^3)$	54
4.2.1 Separable Scaling Function and Approximation Space.	55
4.2.2 Multiresolution Analysis.	56
4.2.3 Orthonormal Wavelet Basis.	59
4.2.4 Discrete Multiresolution Decomposition Algorithm.	60
4.3 A Motion-Oriented Wavelet Multiresolution Analysis for $L_2(\mathbb{R}^3)$	63
4.3.1 Decoupling the Spatial and Temporal Decomposition Process. . .	64
4.3.2 Discrete, Motion-Oriented Decomposition Algorithm.	69
4.4 Spatio-Temporal Aliasing and Fourier Frequency Motion Analysis	75
4.5 Applications and Results	79
4.6 Conclusions	91
V. Object Discrimination Using a Motion-Oriented Wavelet Multiresolution Analysis . .	93
5.1 Introduction	93
5.2 A <i>Vector</i> Wavelet Motion Sensor	93
5.2.1 The Hilbert Transform	95
5.2.2 Directionally Selective Wavelet Filters	97
5.2.3 Directionally Selective, Motion-Oriented Multiresolution Wavelet Analysis	104
5.2.4 Computing Optical Flow	112
5.3 A Cooperative-Competitive Flow Restoration Mechanism	124

	Page
5.3.1 Modified Gated Dipole Filter.	124
5.3.2 Methodology	129
5.3.3 Applications and Results.	135
5.4 Conclusion	139
 VI. Increasing the Speed of the Motion-Oriented Multiresolution Wavelet Decomposition Algorithm through Digital and Optical Parallelization	 144
6.1 Introduction	144
6.2 Digital Parallelization	144
6.2.1 Serial Motion-Oriented Wavelet Decomposition Algorithm . . .	145
6.2.2 Parallel Algorithms for Distributed SUN SPARCstation 2 Network and Intel iPSC/2 and iPSC/860 Hypercubes.	146
6.2.3 Tests and Results	148
6.3 Optical Parallelization	150
6.3.1 Optical Wavelet Theory.	151
6.3.2 System Design	153
6.3.3 Tests and Results	155
6.4 Conclusion	158
 VII. Conclusion and Contributions	 160
7.1 Conclusion	160
7.2 Contributions	161
 Bibliography	 166
 Vita	 170

List of Figures

Figure	Page
1. Spatial and temporal frequency sensitivity of motion cells	3
2. STFT and wavelet filter banks	11
3. Haar scaling function and wavelet	16
4. Daubechies 4 and 8 QMF pairs	17
5. Truncated cubic spline wavelet and its Fourier transform	18
6. 1D wavelet decomposition algorithm and frequency representation	19
7. 2D wavelet decomposition algorithm and frequency representation	21
8. Pictorial representation of four common constraints used in optical flow computations (5).	24
9. Direction ambiguity induced by a moving edge as seen through a circular aperture . .	26
10. Delta pair in spatio-temporal frequency space and lines of constant phase associated with inverse Fourier transform of delta pair	30
11. Stationary and moving 1D rectangles and their Fourier transforms	32
12. Dashed lines outline plane generated in 3D spatio-temporal frequency space by a two dimensional object moving with constant vertical velocity. Circles illustrate shift of a single frequency component.	33
13. Bank of spatio-temporal Gabor filters used by Watson and Ahumada to compute local velocity components (58).	34
14. Oct-tree sub-band coding structure	44
15. Frequency representation of spatio-temporal approximation signal	46
16. Frequency representation of spatio-temporal detail signals	47
17. Simple animated image sequence	48
18. Spatio-temporal detail coefficients of animated scene	49
19. Multiscale results obtained by applying S. Mallat's 2D multiresolution analysis to a simple rectangle	50
20. Conventional and unconventional wavelet filterbanks for discriminating two identical 1D objects moving at different velocities	51

Figure	Page
21. Frequency supports of two 1D objects moving at same speed in opposite directions superimposed on conventional wavelet filterbank	52
22. Non-homogeneous oct-tree sub-band coding structure	64
23. "Decoupled" array of embedded spatio-temporal subspaces	67
24. A visualization of the 3D motion-oriented wavelet decomposition process.	70
25. Spatial, and temporal decomposition algorithms for 3D motion-oriented multiresolution wavelet analysis	70
26. A visualization of the frequency support in the Fourier plane of the basis functions for each space generated by the 3D wavelet multiresolution motion decomposition. . . .	73
27. A visualization of the frequency support in the Fourier plane of the basis functions for each space generated by the 2D wavelet multiresolution motion decomposition. . . .	74
28. Diagonally moving gaussian and its Fourier transform	77
29. Effects of spatial and temporal aliasing on FFT of moving gaussian	79
30. Two rectangles traveling at different speeds and their Fourier transforms	80
31. Frequency supports of the wavelet filters generated by several decompositions in time for a plane taken through the Fourier transform of the image sequence in previous figure at the spatial frequency $f_y = \pi$	81
32. Segmenting two equally sized rectangles moving horizontally with different speeds .	83
33. Segmenting two differently sized rectangles moving horizontally with different speeds	84
34. Frequency supports of the wavelet filters generated by several spatial decompositions for a plane taken through the Fourier transform of the image sequence in previous figure at the temporal frequency $f_t = \pi$	84
35. Segmenting the diagonal and horizontal features of the larger, faster rectangle in previous figure	86
36. Several frames of a sequence of IR images in which a large, slow moving tank executes a 180° turn.	87
37. Single, unprocessed frame of detail coefficients from each of several different motion decompositions in space and time	88
38. Several frames of moving tank image sequence processed with a traditional frame-differencing motion extraction technique and motion-oriented wavelet decomposition	89

Figure	Page
39. Two identical rectangles moving in opposite directions and their Fourier transforms	90
40. Diagonal detail coefficients of image sequence in previous figure obtained by one step in the motion-oriented spatial and temporal decomposition processes	91
41. Fourier transform of 3D wavelet graphically rendered using AFIT ray tracing program	94
42. Motion planes in frequency space and their relative orientation to opposing quadrants and octants in a frequency volume	95
43. Analytic wavelet and its Fourier transform	99
44. Table showing the frequency response of extended wavelet in each of eight octants in a spatio-temporal frequency volume	101
45. Fourier transform of 3D "extended" wavelet graphically rendered using AFIT ray tracing program	102
46. Definition of octants in 3D frequency volume	103
47. Flow diagram depicting one step in a process designed to capture frequency information contained in diagonally opposing frequency regions of a symmetric wavelet filter	108
48. Several frames of 64×64 imagery containing two identical rectangles traveling with the velocity components $(v_x = 1, v_y = 0)$ frame/sec. and $(v_x = 0, v_y = 1)$ frame/sec.	109
49. Coefficients obtained by segmenting horizontally moving object using directionally selective motion-oriented multiresolution wavelet analysis	110
50. Linear supporting region in 2D frequency space of the Fourier transform of a 1D object translating with some velocity v_x	113
51. Digital center frequency triplets of diagonally opposing filters generated by discrete directionally selective motion-oriented multiresolution wavelet analysis	114
52. Several frames of 3D image sequence containing diagonally moving gaussian brightness pattern	115
53. Optical flow map of diagonally moving gaussian brightness pattern	118
54. Two gaussians moving with different speeds and their optical flow	119
55. Multiple gaussians moving with different velocities and their optical flow	120
56. Two gaussian brightness patterns with different variances and velocities	121
57. Multiresolution optical flow	122

Figure	Page
58. Synthetic tank in occluding background and optical flow	123
59. Two views of the right and left hand lobes of a dipole filter lying on a horizontal orientation plane	126
60. A 2D projection of a dipole lobe onto the kth orientation plane	127
61. Major parameters in dipole filter	128
62. Flow completion properties of dipole filter as a function of dipole width	129
63. Outputs of oriented flow maps sum to form interpolated flow vector	130
64. Several examples of consistent and inconsistent flow behavior.	131
65. Key elements of cooperative-competitive flow restoration methodology	132
66. Using dipole and activation threshold to weaken/strengthen flow consistency requirement	134
67. Correlated gaussian noise and resulting flow map with and without flow restoration methodology	137
68. Moving gaussian brightness pattern in correlated gaussian noise	137
69. Flow map of gaussian plus noise after applying flow restoration methodology	138
70. Four image sequences generated by adding equally correlated gaussian noise to synthetic tank image sequence	140
71. "Flow restored" optical flow maps from previous four tank sequences	141
72. A visualization of the motion-oriented multiresolution wavelet decomposition process.	146
73. Major processing tasks associated with the motion-oriented wavelet decomposition algorithm	147
74. Speed-up vs. number of nodes for the iPSC/2 implementation using a Daubechies 8 tap filter in space and time	149
75. Speed-up vs. number of nodes for the iPSC/860 implementation using a Daubechies 8 tap filter in space and time	150
76. Two dimensional Haar mother wavelet and its Fourier transform	153
77. Theoretical results from a digital simulation of the optical continuous Haar wavelet transform	154
78. Optical wavelet transform designs	155
79. Optical Haar wavelet implemented with ternary state MOSLM	156

Figure	Page
80. 16×16 spatially filtered wavelet correlated with binarized input image.	157
81. 32×32 Haar wavelet produced using the aperture stop wavelet design method. . . .	157
82. Comparison of digital and optical results obtained by correlating binarized input image with 8×8 Haar wavelet	158

Abstract

This research presents a multiresolution wavelet analysis tool for analyzing motion in time sequential imagery. A theoretical framework is developed for constructing an $L_2(\mathbb{R}^3)$ wavelet multiresolution analysis from three non-identical spatial and temporal $L_2(\mathbb{R})$ wavelet multiresolution analyses. This framework provides the flexibility to tailor the spatio-temporal frequency characteristics of the three dimensional wavelet filter to match the frequency behavior of the analyzed signal. An unconventional, discrete multiresolution wavelet decomposition algorithm is developed which yields a rich set of independent spatio-temporally oriented frequency channels for analyzing the size and speed characteristics of moving objects. Unlike conventional wavelet decomposition methods, this algorithm provides independent zoom-in and zoom-out capability in space and time. Symmetric 3D filters produced by the unconventional decomposition process are combined with the properties of the Hilbert transform to produce a bank of directionally selective wavelet filters. Multiple directionally selective wavelet filters are integrated to form a multiresolution vector wavelet motion sensor capable of unambiguously computing the optical flow of a 3D image sequence. A unique flow restoration methodology is presented which incorporates a modified version of Grossberg's gated dipole filter in a cooperative-competitive flow restoration methodology that reinforces consistent flow behavior and removes flow inconsistencies. Finally, several digital and optical parallel architectures are investigated for their ability to speed up the 3D wavelet decomposition process.

A NON-HOMOGENEOUS, SPATIO-TEMPORAL,
WAVELET MULTIREOLUTION ANALYSIS
AND ITS APPLICATION TO THE ANALYSIS OF MOTION

1. Introduction

1.1 Historical Background

Modern military target identification systems primarily detect and track *heat sources* in an infrared (IR) image. They often require a "person-in-the-loop" for acquiring a potential target. Once acquired, the target is generally tracked by first thresholding a high contrast scene for "hot spots" and then applying one of several tracking techniques including 2D frame-to-frame feature matching, centroid matching and correlation matching (7, 50). These systems make limited use of *a priori* information, target models and other scene analysis techniques used in the computer vision field.

Computer vision target segmentation and recognition systems generally employ a "static is basic" strategy in which single, static image frames from a time sequence of two dimensional imagery are analyzed individually for attributes such as texture, color and boundaries. The results are then later connected in various ways *across* time (5, 20). However, studies of biological systems show the analysis of information *in* time provides animals with extremely valuable clues for segmenting and identifying moving objects in a dynamic scene (37, 51, 52). Indeed, some animals, such as the frog, employ a "motion is basic" perceptual strategy in which stationary objects, such as a dead fly, are evidently ignored during their internal object segmentation and recognition computations. Additionally, neurophysiological research in higher order mammals (e.g., the macaque) has uncovered anatomically distinct visual pathways devoted exclusively (at least in the early stages of the visual processing system) to motion analysis (43, 61).

From a biological perspective, then, it appears some types of object recognition problems may be better-suited to the analysis of information *in* time (a spatio-temporal process) as opposed to the analysis of information *across* time (a spatial *and* temporal process). Thus, the purpose of this research effort was to focus on the use of motion cues as a means of facilitating the pattern recognition process. In

particular, the research concentrated on one of the earliest steps in the motion-based pattern recognition process: determining the location, speed and direction of objects moving in a scene. The motion-based object discrimination strategy employed in this research draws on the results of biological research into the motion detection properties of the mammalian visual cortex.

The accumulated results of past and ongoing research into the motion detection properties of cortical cells, layers of cells, and the interconnections between cells in the mammalian visual cortex provide four valuable clues for the construction of a computer vision-based motion detection system. First, motion is perceived locally. Humans are able to perceive different motions in different parts of the scene. However, little is known about the size of localized motion detection regions (57). Second, perceived motion is spatial frequency specific. Individual motion sensors tend to respond to a specific band of spatial frequencies (2). The average spatial frequency bandwidth of these cells is approximately one octave. Third, motion sensors are selective for speed. Indeed, cortical motion detection cells can reliably detect speed variations on the order of 5% (41). And fourth, motion detection cells exhibit a spatio-temporal contrast sensitivity that determines the range of spatial frequencies detectable for moving objects (49). A 2D representation of the spatio-temporal contrast sensitivity data collected by Robson is shown in Figure 1. Here, temporal frequency corresponds to the speed at which a horizontally oriented sinusoidal grating moves past a viewer's field of view. Evidently, spatial and temporal frequencies that lie outside the diamond shaped region cannot be detected by the human visual system. These clues clearly point towards the existence of a biological motion detection system in mammals that responds to localized spatial and temporal frequency stimuli.

Historically, motion analysis algorithms have employed frame-to-frame processing techniques such as block matching, feature correspondence and spatio-temporal gradient analysis to characterize objects moving in a 2D scene (5, 26, 27, 55). These techniques require densely sampled imagery in space and time - which make them computationally expensive - and each is highly susceptible to the presence of noise in the spatio-temporal imagery. Additionally, their primary emphasis has been on the construction of a velocity field that depicts movement, rather than on the task of explicitly segmenting moving objects in the scene.

In the mid-to-late 1980s, several researchers began to explore a spatio-temporal frequency motion analysis approach that required the integration of several frames of time sequential imagery (1, 19, 25, 57). Each of these approaches is based on the observation that the Fourier transform of

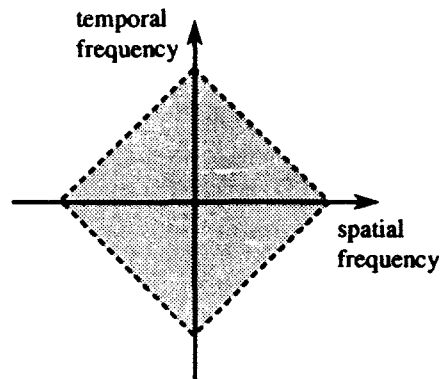


Figure 1. Diamond shaped spatio-temporal frequency plot of Robson's experimental data relating spatial and temporal frequency sensitivity of motion cells in primary visual cortex. Temporal frequency corresponds to the speed at which a horizontally oriented sinusoidal grating moves past a viewer's field of view (49).

a 2D brightness pattern moving with constant velocity across a 2D image plane lies on a plane in spatio-temporal frequency space whose coordinates are governed by the x and y velocity components of the object. In order to determine the orientation of the plane in frequency space, and, therefore, the velocity of the moving object, each of these approaches employ heuristic spatio-temporal filtering techniques that provide little control over inter-dependent filtering characteristics such as filter overlap, filter bandwidth, and space-time/frequency localization. Additionally, these approaches use rigid filter designs that cannot be easily modified to meet a particular problem scenario. Finally, none of the approaches are applied in the presence of noise. This research carries forward the Fourier filtering concepts in the examples cited above, and combines them with the mathematical rigor of the wavelet multiresolution analysis to yield a unique and powerful motion analysis tool that discriminates moving objects in noise-corrupted imagery based on their size, speed and directional properties.

1.2 Problem Statement and Scope

Accurately detecting and discriminating multiple objects moving across a 2D sensor array in the presence of physical and system noise is an unsolved problem. This research studies the feasibility of using a spatio-temporal wavelet multiresolution analysis for this purpose. The analysis of this solution strategy focuses on several key areas, including 1) the extension of existing 2D wavelet multiresolution analysis theory to three dimensions, 2) the creation of separable, non-homogeneous, wavelet filters

that enhance the flexibility of the motion analysis process, 3) decoupling the spatial and temporal multiresolution decomposition processes to provide the ability to independently analyze spatial and temporal details in a 3D signal, 4) the construction of a wavelet filter bank that provides directional selectivity, 5) combining the coefficients obtained in the decomposition process to estimate localized velocity information in the presence of physical and system noise phenomena, and 6) the investigation of several digital and optical parallelization techniques to determine their ability to increase the speed of the 3D wavelet motion analysis process. The research contributions made in these areas are briefly reviewed below.

1. *A Three-Dimensional Wavelet Multiresolution Analysis.* Current wavelet literature focuses on the multiresolution analysis of 1D time signals and 2D images. Y. Meyer's theory provides for the extension of the separable wavelet multiresolution analysis to $\mathbb{R}^n$, however, no details are provided for constructing the corresponding orthonormal wavelet basis set (42). This research shows that each detail space in the 3D multiresolution analysis is spanned by integer translations of a set of seven wavelets, and that the family of wavelets consisting of all possible dyadic dilations of these seven wavelets forms an orthonormal basis for $L_2(\mathbb{R}^3)$. Additionally, an "oct-tree" sub-band coding scheme for implementing a "Discrete Spatio-temporal Wavelet Transform" is developed which generates a bank of non-overlapping octave-band filters with identical - or "homogeneous" - spatial and temporal frequency characteristics.
2. *A Non-Homogeneous Three-Dimensional Wavelet Multiresolution Analysis.* In the conventional multiresolution scheme introduced by Meyer and Mallat, the 2D approximation space V_j was created from two identical 1D approximation spaces. This generates a wavelet filter with identical frequency characteristics in the f_x and f_y spatial frequency dimensions. Similarly, in the 3D "conventional" extension described above, the tensor product of three identical approximation spaces was formed to create a 3D approximation space whose corresponding filter has identical passband characteristics in f_x , f_y and f_t . However, this approach does not provide the flexibility to tailor the spatio-temporal frequency characteristics of the wavelet filter to match the frequency behavior of the 3D signal under analysis. This section proves one can construct a "non-homogeneous" wavelet multiresolution analysis and corresponding orthonormal wavelet basis for $L_2(\mathbb{R}^3)$ from non-identical spatial and temporal filters, thereby increasing the flexibility of the wavelet filter design process.

3. *A Motion-Oriented Multiresolution Wavelet Analysis: Decoupling the Spatial and Temporal Decomposition Processes.* At each stage in the "conventional" non-homogeneous 3D wavelet decomposition algorithm, the spatial and temporal samples of the approximation and detail signals are both equally decimated to yield a bank of analysis filters whose spatial and temporal bandwidths both decrease by a factor of two from one stage of the decomposition to the next. Thus, at any level in the decomposition process, one is required to analyze the signal at equal scales in space and time. It is shown, however, that the analysis of moving objects requires the ability to examine the signal across multiple scales in time for a fixed scale in space. Thus, an unconventional 3D wavelet decomposition theory and algorithm are presented that maintains the orthogonality properties of the analyzing wavelets, employs a sub-band, multirate coding scheme for rapid signal analysis, and allows one to independently zoom-in and zoom-out on spatial and temporal details in the scene.
4. *A Vector Wavelet Motion Sensor.* The motion-oriented multiresolution wavelet analysis described above was designed to detect objects of different sizes moving with different speeds across a two-dimensional image plane. The symmetric 3D filters produced by the decomposition process thus act as a scalar motion sensing detectors in that they respond to the magnitude of an object's velocity vector (i.e., its speed), rather than to the vector quantity of speed and direction. In order to obtain directional selectivity, the independently scaled wavelets are combined with the traditional properties of the Hilbert Transform to yield an orthogonal set of wavelet motion sensors that capture signal energy in diagonally opposing regions of frequency space. The response of these sensors are then combined to compute the localized speed and direction of a moving object over multiple scales in space and time.
5. *A Cooperative-Competitive Optical Flow Restoration Mechanism.* The performance of the wavelet-based flow estimation algorithm developed under this research effort is degraded by the presence of physical and system noise phenomena. Therefore, a unique flow restoration methodology is presented that incorporates a modified version of S. Grossberg's gated dipole filter in a cooperative-competitive flow restoration methodology that reinforces consistent flow behavior and removes flow inconsistencies. The vector wavelet motion sensor is then used in conjunction with the cooperative-competitive flow restoration algorithm to discriminate moving objects in noise-corrupted 3D imagery.

6. *Digital and Optical Parallelization Techniques for Increasing the Speed of the Motion-Oriented Wavelet Decomposition Algorithm.* The bulk of the processing time required to run the wavelet vector motion analysis algorithm is taken up by the motion-oriented 3D wavelet decomposition process. Thus, several digital and optical parallel architectures are investigated to determine their potential for increasing the computational speed of the motion oriented decomposition algorithm. The digital parallel algorithms were implemented on a distributed SUN SPARCstation 2 network, an Intel iPSC/2 Hypercube, and an iPSC/860 Hypercube. The optical architectures employ a SEMETEX 128×128 Magneto-optic Spatial Light Modulator and thermo-plastic holography to implement the 2D spatial decomposition stage of the 3D motion-oriented wavelet decomposition algorithm.

1.3 Dissertation Organization

This dissertation is organized into seven main chapters. The following chapter presents background material that serves as a foundation for this research. The concepts of a continuous wavelet transform and a wavelet multiresolution analysis are reviewed, including the non-uniform filter bank properties of the conventional 2D wavelet multiresolution analysis. A cursory description of several techniques used to compute an optical flow field are then reviewed, followed by a deeper examination of existing methods for computing the optical flow from spatio-temporal frequency information. Chapter III describes the extension of the conventional 2D multiresolution analysis to three dimensions. It is shown that the detail space between two spatio-temporal approximations spaces is spanned by integer translations of seven wavelets. In order to enhance the flexibility of the 3D wavelet design process, Chapter IV proves one can construct a spatio-temporal multiresolution analysis by forming the tensor product of three non-homogeneous 1D scaling functions. The theory is also developed for an unconventional multiresolution wavelet analysis that allows one to independently control the spatial and temporal analysis levels in the decomposition process. In Chapter V, the unconventional decomposition technique is combined with the properties of the Hilbert transform to generate a wavelet vector motion analysis tool that is selective for the size, speed and direction of objects moving in a 2D image plane. The vector motion tool is combined with a non-linear, competitive-cooperative flow enhancement technique that provides the ability to compute the optical flow in the presence of system and physical noise phenomena. Chapter VI then presents several parallel digital and optical architectures

designed to increase the speed of the 3D motion-oriented multiresolution decomposition algorithm. The final chapter of the document provides a brief conclusion and lists the individual contributions made throughout this research effort.

II. Background Material

2.1 Introduction

This research fuses the concept of motion analysis using spatio-temporal frequency (STF) information, with the analytical capabilities of a 3D wavelet multiresolution analysis. Many of the theoretical contributions made here lie in the broadly defined area of wavelet transform theory. Thus, this chapter provides a brief overview (as opposed to a rigorous mathematical analysis) of some important wavelet related concepts. These concepts include the continuous wavelet transform and its relationship to a multi-scale correlation process, the wavelet series approximation of a real, finite energy signal, and the wavelet multiresolution analysis as first introduced by S. Mallat and Y. Meyer (39, 40). The second major section of this chapter reviews the advantages and limitations of several non-STF motion characterization techniques. It should be noted here that although several "motion characterization" algorithms exist for the purpose of computing 3D structure from kinematic motion data (for a survey of many of these techniques see T. Huang (29)), this research will restrict the concept of "motion characterization" to the problem of assigning a velocity vector to each location in a changing 2D scene. This section also contains a brief discussion of a fundamental problem inherent in any optical flow computation - the aperture problem. The third and final section in the chapter draws a connection between motion and its representation in spatio-temporal Fourier frequency space, and discusses STF motion analysis techniques that apply to this research.

2.2 Signal Analysis with a Wavelet Transform

Wavelet transform theory and, in particular, multiresolution wavelet analyses are gaining popularity in the signal processing community for three main reasons (48). First, they yield orthonormal building blocks for finite energy functions which are considerably more diverse than the complex exponentials found in conventional Fourier analysis. Second each building block has a localized region of support in $\mathbb{R}^n$, making it possible to isolate rapid signal fluctuations over small regions of space or time. And third, the sub-band coding scheme used in discrete multiresolution wavelet decomposition and reconstruction algorithms provides a "fast" method for analyzing and synthesizing signals. This section reviews several key concepts associated with wavelet transform theory that directly apply to the spatio-temporal signal analysis conducted during this research.

2.2.1 *The Continuous Wavelet Transform.* The general definition of a continuous wavelet transform is given in the literature by

$$[W_\psi f](a, b) = \frac{1}{|\sqrt{a}|} \int_{-\infty}^{\infty} f(x) \psi\left(\frac{x-b}{a}\right) dx \quad (1)$$

where f belongs to the vector space $L_2(\mathbb{R})$, ψ is a “wavelet” kernel, and $a, b \in \mathbb{R}$ (although, for practical purposes, the dilation parameter a is typically taken to be greater than zero). In order to reconstruct f from $W_\psi f$, the so-called “admissibility condition” requires that the constant C_ψ must be finite where

$$C_\psi = \int_{-\infty}^{\infty} \frac{|\hat{\psi}(f_x)|^2}{|f_x|} df_x \quad (2)$$

and $\hat{\psi}(f_x) = [\mathcal{F}\psi](f_x)$ is the Fourier transform of the wavelet $\psi(x)$. The admissibility condition ensures ψ decays sufficiently fast to zero at $\pm\infty$. Additionally, to be a true window function (11), it follows from the admissibility condition that the wavelet transform kernel ψ must be “zero mean” in the sense that

$$\int_{-\infty}^{\infty} \psi(x) dx = 0 \quad (3)$$

Taken together, the above conditions on the wavelet kernel imply the graph of ψ must look like a small wave, or *wavelet*.

The continuous wavelet transform is obtained by integrating the signal over all possible shifts and dilations of the wavelet kernel ψ . Through a simple variable substitution, this operation can also be implemented as a correlation process. For example, let the one dimensional spatial wavelet transform, $[W_\psi s](a, b)$, of a signal, $s(x)$, be given by

$$[W_\psi s](a, b) = \int_{-\infty}^{+\infty} s(x) \frac{1}{\sqrt{a}} \psi\left(\frac{x-b}{a}\right) dx \quad (4)$$

where a is a dilation parameter and b is a translation parameter. Letting $\psi_a(x) = \frac{1}{\sqrt{a}} \psi\left(\frac{x}{a}\right)$, Equation 4 can be rewritten as

$$[W_\psi s](a, b) = \int_{-\infty}^{+\infty} s(x) \psi_a(x-b) dx \quad (5)$$

Equation 5 shows that the wavelet transform can be expressed as a correlation process in which the signal is correlated with a scaled and dilated version, ψ_a , of the wavelet ψ . This relationship can

also be described as a filtering operation in the spatial frequency domain. If $S(f_x) = \mathcal{F}\{s(x)\}$, $\Psi(f_x) = \mathcal{F}\{\psi(x)\}$, and $\Psi_a(f_x) = \mathcal{F}\{\psi_a(x)\}$, then Equation 5 becomes

$$[W_\psi s](a, b) = [\mathcal{F}^{-1}\{S(f_x) \cdot \sqrt{a}\Psi(-af_x)\}](b) \quad (6)$$

where $\sqrt{a}\Psi(af_x) = \Psi_a(f_x)$.

In many signal analysis applications, including the problem of segmenting and characterizing moving objects based on local spatio-temporal frequency measurements, one must simultaneously analyze the space-frequency or time-frequency behavior of a signal. The most commonly used tool for space-frequency analysis is the Short Time Fourier Transform (STFT). Introduced in its original form by D. Gabor in the 1940s (18), the STFT can be described by the relationship

$$\text{STFT}(\xi, f_x) = \int_{-\infty}^{\infty} f(x)w(x - \xi)e^{-i2\pi f_x x} dx \quad (7)$$

where $w(x)$ is a window function with limited extent and the signal's frequency characteristics are assumed to remain stationary over the width of the window. Equation 7 shows the STFT simply computes the Fourier transform of the portion of the signal enclosed by the window centered at ξ .

One can also view the STFT from a filter bank perspective by considering the frequency behavior of the product $f(x)w(x - \xi)$. From Fourier transform theory, the frequency spectrum of the product is obtained by convolving the Fourier transform, $F(f_x)$, of $f(x)$ with the Fourier transform, $e^{-i2\pi\xi f_x}W(f_x)$, of $w(x - \xi)$. This yields the following alternative expression for Equation 7

$$\begin{aligned} \text{STFT}(\xi, f_x) &= F(f_x) * W(f_x)e^{-i2\pi\xi f_x} \\ &= \int_{-\infty}^{\infty} F(q)W(f_x - q)e^{-i2\pi\xi(f_x - q)} dq \end{aligned} \quad (8)$$

where $*$ indicates convolution and q is a dummy variable. If one defines the transfer function $H(q)$ by $H(q) = W(q)e^{-i2\pi\xi q}$, and if the analysis is restricted to a single frequency, say f_o , then the above integral becomes:

$$\text{STFT}(\xi, f_o) = \int_{-\infty}^{\infty} F(q)H(f_o - q) dq \quad (9)$$

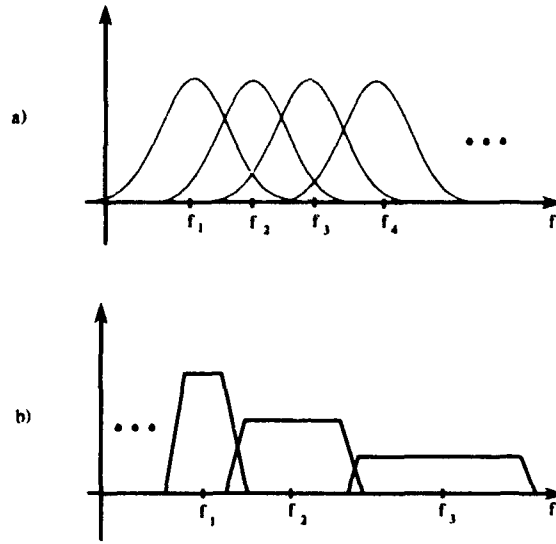


Figure 2. a) The STFT represented as a collection of uniform filters centered around a discrete set of frequencies. b) Representing the continuous wavelet transform as a bank of non-uniform filters with logarithmic coverage.

Equation 9 shows that when evaluated at a single frequency, the STFT can be expressed as a windowing operation in frequency space where the window location is f_0 and the size of the window is determined by the bandwidth of the transfer function H . Thus, Equation 9 represents the Fourier frequency components of f localized in space around ξ and localized in frequency around f_0 . If the operation is repeated for many discrete frequency values, the analysis amounts to a uniform filter bank representation of the signal as shown in Figure 2a (56).

The major drawback to the STFT as a time-frequency analysis tool is that once a window is chosen, the space-frequency resolution remains fixed over all space and all frequencies. This implies the STFT analyzes long duration, low frequency components and short-duration, high frequency components with the same window, which can lead to inaccurate estimates of the location and frequency content of both signal types (56). This problem is of particular concern when attempting to analyze the motion of multiple objects in a scene, each of which may have a different size and velocity. The continuous wavelet transform overcomes this problem by replacing the fixed width window with a prototype, or “mother”, wavelet where all impulse responses of the filter bank and their Fourier

transforms are scaled versions of the mother, i.e.,

$$\frac{1}{\sqrt{a}}\psi\left(\frac{x}{a}\right) \Leftrightarrow \sqrt{a}\Psi(af) \quad (10)$$

Thus, unlike the STFT in Equation 9 where all the responses are obtained by a frequency *shift*, the responses of the continuous wavelet transform are obtained by a frequency *scaling* operation. Furthermore, if one constrains the dilation parameter of the wavelet so that a constant ratio is maintained between the bandwidth, Δf , and center frequency, f_c , of the filters associated with the impulse responses (i.e., $\frac{\Delta f}{f_c} = c$), then the filter bank representation of the wavelet transform consists of non-uniform filters spread logarithmically over the frequency axis (Figure 2b)). Moving out along the frequency axis in Figure 2b), the bandwidth of each filter increases by an octave, implying the spatial width of the corresponding wavelet impulse response decreases by an octave. The advantage of this type of "constant Q" filter bank approach is that spatial resolution becomes arbitrarily good at high frequencies, and frequency resolution becomes arbitrarily good at low frequencies. Thus, the continuous wavelet transform can eventually resolve two narrowly separated spatial impulses simply by increasing the analyzing frequency (i.e., reducing the analyzing scale) until the spatial dilation of the corresponding wavelet is sufficiently small to separate the two impulses (48).

2.2.2 Signal Approximation With the Wavelet Series. The continuous wavelet transform provides a valuable tool for analyzing the space-frequency behavior of a continuous signal. Another valuable aspect of wavelet theory involves the approximation of finite energy (L_2) signals with the wavelet series (sometimes called the Discrete Wavelet Transform (56)). Like the conventional Fourier series, the wavelet series expands a signal as a weighted superposition of basis elements. However, unlike the Fourier series, which requires signal periodicity and whose complex exponential basis elements cover the entire real line, the wavelet series can be used to represent any L_2 signal by shifted and dilated versions of a prototype wavelet with limited extent.

For example, let $L_2(0, T)$ denote the vector space of T -periodic functions such that

$$\int_0^T |f(x)|^2 dx < \infty. \quad (11)$$

Then any function f in $L_2(0, T)$ has a Fourier series representation given by

$$f(x) = \sum_{n=-\infty}^{\infty} c_n e^{i2\pi \frac{n}{T} x} \quad (12)$$

where the Fourier coefficients are given by the inner product

$$c_n = \frac{1}{T} \int_0^T f(x) e^{-i2\pi \frac{n}{T} x} dx. \quad (13)$$

If $b_n(x) = e^{i2\pi \frac{n}{T} x}$, then the set $\{b_n \mid n \in \mathbb{Z}\}$, forms an orthonormal basis for $L_2(0, T)$. Furthermore, if $b(x) = e^{i2\pi x}$ represents the prototype function for the basis set b_n , then every function in $L_2(0, T)$ is obtained by a superposition of *dilations* of the prototype function.

Now consider the vector space $L_2(\mathbb{R})$ where $f \in L_2(\mathbb{R})$ implies

$$\int_{-\infty}^{\infty} |f(x)|^2 dx < \infty \quad (14)$$

Clearly the set of complex sinusoidal functions $b_n(x)$ can no longer serve as a basis set for $L_2(\mathbb{R})$, since they don't belong to $L_2(\mathbb{R})$. Additionally, since each vector in $L_2(\mathbb{R})$ decays to zero at $\pm\infty$, it seems reasonable that candidate functions for an $L_2(\mathbb{R})$ basis set must themselves decay rapidly to zero. Finally, for practical purposes, it is advantageous to generate each element in the basis set by scaling and shifting a single prototype function as in the case of the wavelet kernel for the continuous wavelet transform (14). With these considerations in mind, one can define the coefficients of the wavelet series by (56)

$$c_{j,n} = \int_{-\infty}^{\infty} f(x) a_o^{\frac{j}{2}} \psi(a_o^j x - nT_x) dx \quad (15)$$

where the sampling parameters a_o and T_x are constants and $n \in \mathbb{Z}$. The corresponding wavelet series approximation of the L_2 signal f is then given by

$$f(x) \approx \sum_{j=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} c_{j,n} \psi_{j,n}(x) \quad (16)$$

where $j \in \mathbb{Z}$ and $\psi_{j,n}(x) = a_o^{\frac{j}{2}} \psi(a_o^j x - nT_x)$. Notice that the coefficients are obtained by discretizing the dilation and shift parameters of the continuous wavelet transform. For this reason, Equation 15 is often referred to as the Discrete Wavelet Transform, in which case the wavelet series approximation

in Equation 16 becomes the means by which the signal f is reconstructed from its Discrete Wavelet Transform. The idea, then, is to choose values of a_o, T_x and the mother wavelet $\psi(x)$ so that the wavelet series approximation is close (under the L_2 norm) to the signal f using as few coefficients as possible. (11).

From the standpoint of this research, there are two important points to make about Equations 15 and 16. First, generally speaking, the dilated and shifted kernel in the discrete wavelet transform integral (Equation 15) and the basis elements in the series approximation (Equation 16) are not necessarily the same function. In order to form a wavelet series approximation, it is only required that the kernel and corresponding prototype basis element are duals of one another (11). Throughout this research, however, only identical kernels and wavelet basis elements will be considered. Second, the set $\{\psi_{j,k}\}$ is not necessarily an orthonormal basis for $L_2(\mathbb{R})$. It must simply be a "stable" basis, which may introduce redundancy between coefficients in the approximation (11). In general terms, the amount of redundancy in the series approximation is determined by the sampling parameter a_o . If a_o is set equal to two and T is chosen to be one, as will be the case with this research, then, under special circumstances for the choice of ψ , the wavelet basis set will be orthonormal (13). Assuming this condition holds, the next section addresses how one constructs such an orthonormal basis set using a wavelet Multiresolution Analysis.

2.2.3 The Wavelet Multiresolution Analysis in $\mathbb{R}^1$ and $\mathbb{R}^2$. A multiresolution analysis consists of a chain of closed linear spaces V_j which satisfy (11)

$$\cdots V_{-2} \subset V_{-1} \subset V_0 \subset V_1 \subset V_2 \subset \cdots \quad (17)$$

where

$$\overline{\bigcup_{j \in \mathbb{Z}} V_j} = L_2(\mathbb{R}); \quad \bigcap_{j \in \mathbb{Z}} V_j = \{0\} \quad (18)$$

and

$$\begin{aligned} f(x) \in V_j &\Leftrightarrow f(2x) \in V_{j+1}; \quad j \in \mathbb{Z} \\ f(x) \in V_j &\Rightarrow f\left(x + \frac{n}{2^j}\right) \in V_j; \quad n \in \mathbb{Z} \end{aligned} \quad (19)$$

S. Mallat has shown that if the chain of subspaces in Equation 17 meets these requirements, then there exists a unique "scaling" function $\phi(x) \in L_2(\mathbb{R})$ such that $\{2^{\frac{j}{2}}\phi(2^j x - n) \mid n \in \mathbb{Z}\}$ is an orthonormal basis for V_j (often referred to as an *approximation space*) (39). Furthermore, denoting the orthogonal complement of V_j in V_{j+1} by W_j where

$$V_{j+1} = V_j \oplus W_j \quad (20)$$

one can create a mother wavelet $\psi(x)$ such that $\{2^{\frac{j}{2}}\psi(2^j x - n) \mid n \in \mathbb{Z}\}$ is an orthonormal basis for W_j (here $\oplus$ indicates the direct sum). The spaces W_j where $j \in \mathbb{Z}$ are mutually orthogonal; thus, by the denseness property of the multiresolution analysis the set of scaled and dilated wavelets $\{2^{\frac{j}{2}}\psi(2^j x - n) \mid (j, n) \in \mathbb{Z}^2\}$ forms an orthonormal basis for $L_2(\mathbb{R})$. The scaling functions and the mother wavelet are related by the "two-scale" recursion relations

$$\begin{aligned} \phi(x) &= \sum_{n=-\infty}^{\infty} h_n \sqrt{2} \phi(2x - n) \\ \psi(x) &= \sum_{n=-\infty}^{\infty} g_n \sqrt{2} \phi(2x - n) \end{aligned} \quad (21)$$

where the coefficients h_n and g_n are discussed below. W_j is typically referred to as the j th *detail space*, because it captures the difference in signal information between the approximation spaces V_{j+1} and V_j .

Approximation and detail signals are created by orthogonally projecting the input signal onto the appropriate approximation or detail space. Since each space is spanned by an orthonormal basis set, the signal projection onto a given approximation or detail space at, say, the j th resolution, is equivalent (i.e., isometrically isomorphic) to the sequence of projection coefficients obtained by the inner product operations

$$\begin{aligned} a_{j,n} &= \int_{-\infty}^{\infty} f(x) 2^{\frac{j}{2}} \phi(2^j x - n) dx \\ d_{j,n} &= \int_{-\infty}^{\infty} f(x) 2^{\frac{j}{2}} \psi(2^j x - n) dx \end{aligned} \quad (22)$$

where $a_{j,n}$ and $d_{j,n}$ represent the j th approximation and detail coefficients respectively. The coefficients in Equation 22 are obtained through a convolution operation in which the output is sampled at the discrete

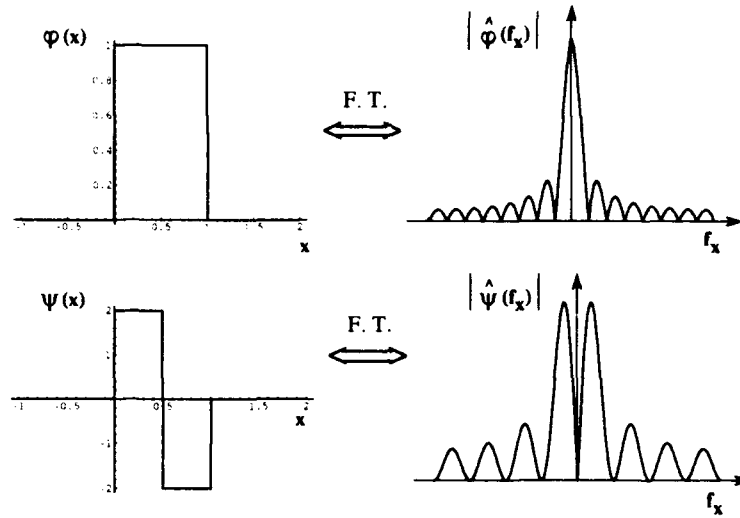


Figure 3. Top: Haar scaling function and its Fourier Transform. Bottom: corresponding Haar wavelet and its Fourier Transform.

points $\frac{k}{2^j}$, $k \in \mathbb{Z}$. Thus, based on earlier discussions regarding the continuous wavelet transform, one might intuitively conclude that the projection operations onto the approximation and detail spaces can be represented by low and bandpass filtering operations, where the width of the filters depend on the dyadic scale 2^j of the scaling and wavelet functions. This is indeed the case as demonstrated by the frequency behavior of the Haar scaling function and wavelet contained in Figure 3 (39).

In practice, the wavelet multiresolution analysis is implemented with a pyramidal sub-band coding scheme introduced by Mallat (39). Following Mallat's approach, a discretely sampled version of an $L_2(\mathbb{R})$ function, $f(n)$, is projected onto the detail space W_j by capturing the difference in information between orthogonal projections onto the approximation spaces V_{j+1} and V_j . In this scheme, the signal projections are represented by their respective projection coefficients; thus, the algorithm is said to generate a Discrete-Space (or Time) Wavelet Transform (56). The approximation and detail coefficients associated with V_j and W_j (Equation 22) are generated from the approximation coefficients at the next higher scale, V_{j+1} , using a Quadrature Mirror Filter (QMF) pair with impulse responses h_n and g_n and a decimation-by-2 subsampling process. The impulse responses h_n and g_n represent the coefficients in the two-scale relationships defined in Equation 21. Although there are several possible ways to define the relationship between the impulse responses (14, 48), the relationship

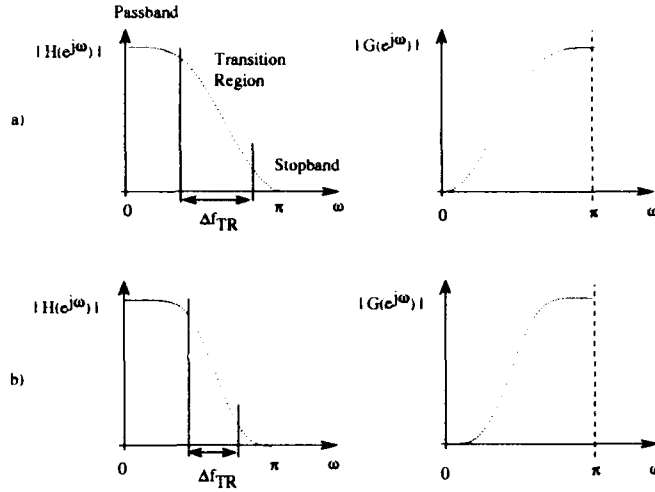


Figure 4. The magnitudes of the frequency responses of the Daubechies order a) 4 and b) 8 QMF pair. Note how the transition region decreases as the filter order increases.

used throughout this research, is given by $g_n = (-1)^{1-n}h_{1-n}$, where h_n is formed by computing the inner product between $\phi(\frac{u}{2})$ and $\phi(u - n)$.

Perhaps the most commonly used QMF pairs in the wavelet literature are those constructed by I. Daubechies (14). Throughout this dissertation, a Daubechies filter pair will be referred to as "Daubechies N " where N is the number of coefficients in the impulse response of the filter. Daubechies' filter pairs are easy to implement digitally because they have a finite impulse response (FIR). Additionally, the transition region, Δf_{TR} , between a Daubechies filter's passband and stopband narrows as the order of the filter increases as shown by the frequency responses of the Daubechies 4 and Daubechies 8 QMF pairs in Figure 4. Since the h and g filter coefficients are tabulated for a large number of filter orders (13), this allows one to easily change the filter's cutoff frequency characteristics to meet a given design constraint. Unfortunately, there is a major drawback to using Daubechies' QMF pairs for the type of image processing work done in this research effort: their frequency responses do not have linear phase.

A FIR filter has linear phase if and only if its impulse response is symmetric or antisymmetric, i.e., $h(n) = \pm h(N - n)$ where N is the order of the filter (56). Daubechies QMF pairs are asymmetric, thus they do not have linear phase. This can pose serious problems when the filters are used in image processing applications (14, 16, 36). Edges and lines in an image are constructed from a sum of

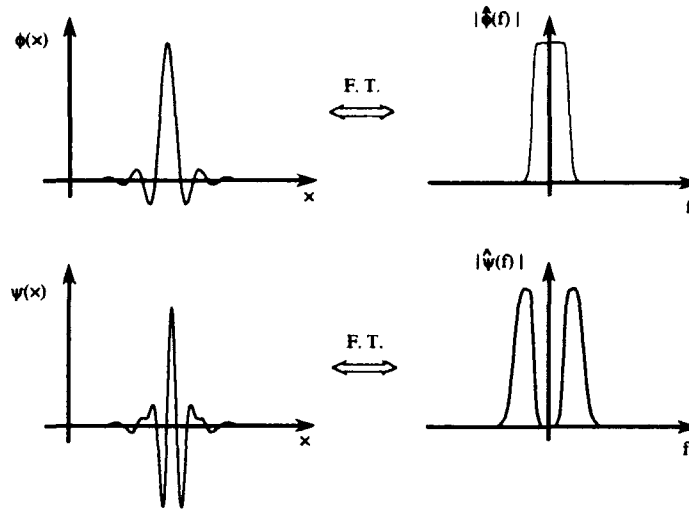


Figure 5. A cubic spline (top) and its corresponding wavelet (bottom), along with their Fourier transforms. Both functions were generated recursively after truncating their impulse responses $h(n)$ and $g(n)$ to 23 coefficients. Note the passband ripple in the Fourier transforms of both functions as a result of the truncation process.

critically aligned 2D frequency components. A high-pass filter with a non-linear phase response can unevenly disperse the frequency components that comprise the edge, causing a blurring effect that reduces the quality of the high-pass filtering operation. Since it was not the purpose of this research to develop a FIR, linear phase QMF design technique, when necessary, symmetric FIR filter pairs were constructed by equally truncating the h and g impulse responses of a symmetric, IIR cubic-spline filter pair. Examples of a symmetric cubic spline scaling function and its corresponding wavelet are shown in Figure 5. The scaling and wavelet functions were constructed recursively from their respective impulse responses after truncating the impulse responses each to 23 coefficients. Note that even after truncation, the DC component of the wavelet remains approximately zero (i.e., the wavelet meets the admissibility condition mentioned earlier). Also note the passband ripple in the Fourier transforms of both functions as a result of the truncation process. A rigorous development of the theory and construction of cubic spline QMF pairs can be found in Chui (11).

A binary tree structure for implementing Mallat's 1D wavelet multiresolution analysis is shown in Figure 6a). The binary tree serves as a "canonical" structure for extending the conventional algorithm to multiple dimensions. In Mallat's pyramidal coding scheme, the coefficients of the $j + 1$ st approximation level are simultaneously decomposed into the j th detail and approximation coefficients using the low-

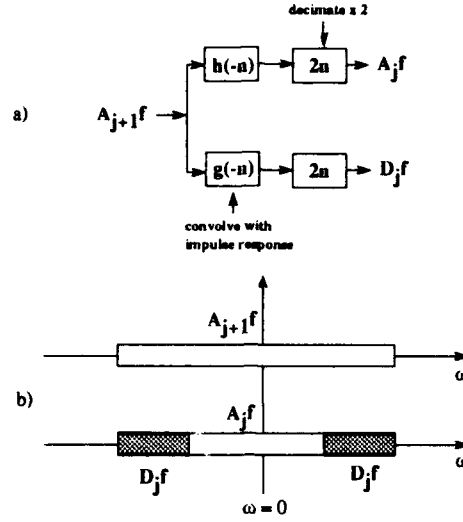


Figure 6. 1D sub-band coding algorithm for decomposing the coefficients of the $j+1$ st approximation level into the coefficients of the j th detail and approximation levels. b) Regions of support along the frequency axis of the approximation and detail signals.

pass and high-pass impulse responses $h(n)$ and $g(n)$. The regions of support in frequency space of the resulting approximation and detail signals are shown in Figure 6b). By repeatedly convolving each approximation signal with $h(n)$ and $g(n)$ and decimating the outputs by a factor of two, the signal is decomposed into frequency bands whose bandwidths and center frequencies vary by octaves. In the signal processing literature, the set of filters generated by multiple stages of the the pyramidal decomposition algorithm is referred to as a two channel paraunitary QMF filter bank (48, 56).

One can also construct a separable orthonormal wavelet basis set for $L_2(\mathbb{R}^2)$ from the chain of "2D" multiresolution approximation spaces $\{V_j \mid j \in \mathbb{Z}\}$, where V_j is defined by (13)

$$V_j = V_j^x \otimes V_j^y = \overline{\text{Span}\{F(x, y) = f(x)g(y) \mid f \in V_j^x, g \in V_j^y\}} \quad (23)$$

$$F(x, y) \in V_j \Leftrightarrow F(2x, 2y) \in V_{j+1} \quad (24)$$

where V_j^x and V_j^y are identical "1D" approximation spaces (i.e., they are spanned by the same scaling function). Here, the 2D scaling function for V_j is formed from the product of both identical 1D scaling functions, and the wavelet orthonormal basis for the orthogonal complement W_j is given by three

wavelets

$$\begin{aligned}\Psi_j^1(x, y) &= 2^j \phi(2^j x) \psi(2^j y) \\ \Psi_j^2(x, y) &= 2^j \psi(2^j x) \phi(2^j y) \\ \Psi_j^3(x, y) &= 2^j \psi(2^j x) \psi(2^j y)\end{aligned}\tag{25}$$

The family of wavelets

$$\{\Psi_j^p(x - m, y - n) \mid j \in \mathbb{Z}; (m, n) \in \mathbb{Z}^2; p = 1, 2, 3\}\tag{26}$$

then forms an orthonormal basis set for $L_2(\mathbb{R}^2)$. Through a straightforward extension of the 1D binary tree structure, one obtains the 2D “quad tree” wavelet multiresolution decomposition algorithm as shown in Figure 7a). Here $A_j f$ and $D_j^n f$, $n = 1, 2, 3$ denote the projection of the $L_2(\mathbb{R}^2)$ image, f , onto the approximation space V_j and detail spaces, W_j^1 , W_j^2 , and W_j^3 spanned respectively by the wavelets $\{\Psi_j^p \mid p = 1, 2, 3\}$ (40). Figure 7b) shows the frequency support of the separable approximation and detail filters used to decompose the 2D image approximation $A_{j+1} f$ into the approximation $A_j f$ and the details D_j^1 , D_j^2 and D_j^3 . Notice that the 2D wavelet decomposition process can be interpreted as a “signal decomposition in a set of independent, *spatially oriented* frequency channels” (40). In Chapter III, the 2D wavelet multiresolution analysis is extended to three dimensions to produce a new signal decomposition tool in which a set of *spatio-temporally oriented* frequency channels are used to analyze movement in 3D imagery. In the following section, several traditional methods for performing this task are reviewed and compared.

2.3 Traditional Methods for Computing Optical Flow

Optical flow is the apparent motion of three-dimensional objects as represented by changing brightness patterns in a 2D image plane. Algorithms designed to compute optical flow attempt to assign a velocity vector to each point in a sampled 2D image plane based on the apparent speed and direction with which brightness patterns move across the image plane. Because optical flow is deduced from the gray-scale content of an image, rather than directly sensed, optical flow computations are often highly susceptible to pixel scintillations caused by noise sources in the imaging process.

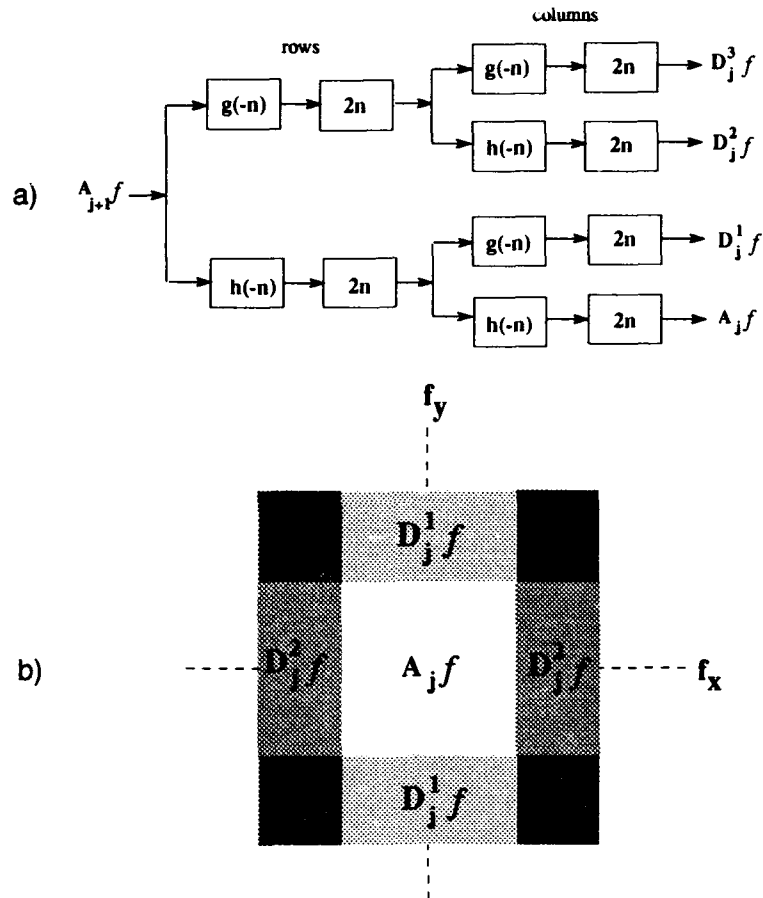


Figure 7. a) Stephane Mallat's 2D discrete multiresolution decomposition algorithm. b) Frequency support of the 2D decomposition of the approximation image $A_{j+1}f$ into $A_j f$ and the detail images $D_j^p f$; $p = 1, 2, 3$ (40).

Optical flow representations are used for a wide variety of applications. Humans, for example, may generate an internal optical flow map as evidenced by our ability to segment moving objects in a monocularly viewed random dot field in the absence of non-motion cues, (30). Optical flow velocity fields are also used to solve the so-called reverse optics problem, where one must deduce the 3D structure and/or motion of real world objects based on changes in their 2D image projections (28, 4, 38). Additionally, some interpolative and predictive video compression schemes use velocity fields to reduce the transmission overhead of television signals (29, 32). More recently, researchers have begun to examine the use of optical flow as a means of segmenting objects in changing 2D imagery (3, 44). Finally, optical flow has been used in automated image understanding algorithms to determine navigational parameters such as time-to-adjacency and time-to-collision (5).

Optical flow computational methods are generally divided into three categories. The first method, often referred to as the correspondence technique, attempts to match blocks of data from one time frame to the next. These blocks may contain gray-scale intensity values (point correspondence) or they may consist of pre-extracted features such as edges or corners (feature correspondence) (33). Correspondence techniques typically attempt to minimize an energy measure that depends on block locations in two subsequent image frames. The second method computes the velocity field by measuring the spatial and temporal intensity gradients surrounding each point in a changing 2D image (3, 4, 26, 27). These measurements are ambiguous in that they produce one equation at each point in the image which must be solved for the velocity components in the x and y directions. Therefore, spatio-temporal gradient techniques must further constrain the velocity field to generate a second equation. Typically, the constrained problem is solved using a variety of optimization techniques. The third method employs spatio-temporal Fourier phase and frequency information to compute optical flow. Fourier *phase*-based computations rely on the well known Fourier transform property that converts shifts in the spatial domain to linear phase terms in the frequency domain (23, 31, 33). Since digital phase is phase wrapped between $-\pi$ and π , these techniques require phase unwrapping routines to compute the velocity of an object whose movement generates phase shifts greater than 2π . Finally, Fourier *frequency* methods compute optical flow using spatio-temporal frequency information (1, 19, 25, 58). Various approaches that fall into each of the above categories are reviewed in the following sections.

2.3.1 Feature Correspondence. Historically, the computation of a 2D velocity flow field has been considered to be a correspondence task (33). That is, after determining the locations of a corresponding pair of features in two subsequent time frames, a displacement vector is assigned to the spatial coordinates underlying these features. The set of features used in the computation vary from a block of gray-scale intensity values, to a complex arrangement of edges, corners, textures, or colors. Features can be extracted using a feature matching template in a pre-processing stage, or, under a more general scheme, they can be chosen arbitrarily by capturing data that exceeds a highly localized contrast detection threshold (33). In any case, the features and their respective spatial coordinates in one time frame are compared with a feature list from the subsequent time frame to find the best match. The spatial displacement between the features is then divided by the interframe time interval to obtain the optical flow for the points in the scene that correspond to the moving features.

Another common optical flow computational method that falls loosely into the category of feature correspondence is called block-matching. This technique is often used in motion-compensated video coding schemes. "Features" in this case are actually gray-scale intensity values contained in a block of pixels (i.e. a pel) of some pre-defined size. Each pel in a time frame is compared with all pels in the next time frame to locate the point that provides the best match between corresponding pel pairs. After assigning a velocity vector to every sample point in, say, a 2D video image array, one can theoretically reduce the transmission overhead of the video imagery by transmitting the original image frame once, and updating it at the decoder with a subsequent set of block-matched displacement vectors.

The measuring stick with which features or blocks are matched between frames varies; however, in most cases, the corresponding feature locations are required to minimize some cost function. A simple example of a standard block matching cost function minimization scheme is one which searches for the optimum pair of spatial coordinates (d_1, d_2) that minimize

$$\mathcal{E} = \sum_{(n,m) \in B_b} [b_p(n - d_1, m - d_2) - b_c(n, m)]^2 \quad (27)$$

where b_c and b_p are blocks of equal dimension in the current and previous frames, and B_b represents the compact region of support of the block (23). In order to increase the efficiency of the matching algorithm, more sophisticated feature correspondence methods often take into account local and/or global motion constraints.

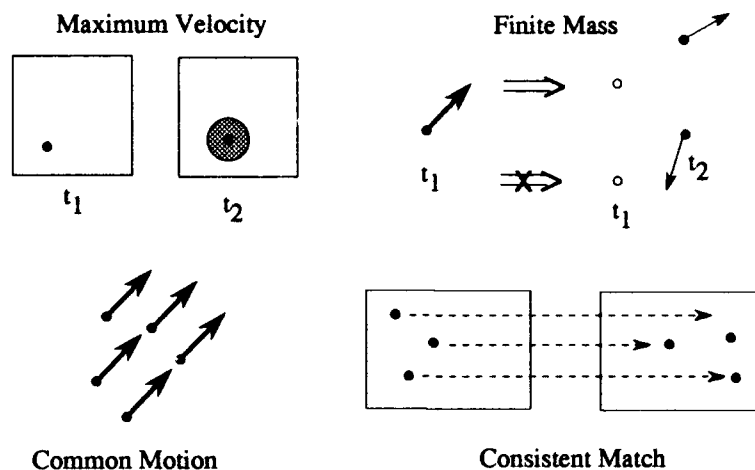


Figure 8. Pictorial representation of four common constraints used in optical flow computations (5).

Ballard lists several of the more common motion correspondence constraints for image frames separated by small time intervals. These are depicted in Figure 8. The maximum velocity constraint assumes a maximum velocity v_{max} for each object in the scene. The maximum displacement of any object then is simply $v_{max}\Delta t$ where Δt is the interframe time interval. Knowing the maximum displacement allows one to limit the best-match search space. The second heuristic is based on the laws of physics which preclude the velocity of objects with finite mass from making discontinuous changes over small time intervals. The third constraint assumes that rigid objects exhibit common motion between frames. The final constraint assumes that two points from one image cannot match a single point from the next image. The primary problem with each of these constraints is that they all assume motion “sinks” and “sources” are absent from the scene (5). This assumption tends to increase the sensitivity of the algorithms to natural phenomena found in military imagery such as noise and occlusions. It is shown in Chapter V that the wavelet-based optical flow algorithm developed in this research functions well despite the presence of these phenomena.

2.3.2 Spatio-Temporal Differentiation. The spatio-temporal gradient approach to computing optical flow measures the first order spatial and temporal gradients around each spatial coordinate in a sequence of densely sampled imagery. These measurements yield the component of motion in the direction of maximally increasing intensity. The motion component that lies perpendicular to the spatial gradient is then determined using a variety of constrained optimization techniques. In this section, the

spatio-temporal gradient optical flow algorithm as originally developed by Horn *et. al.* is presented, along with a brief explanation of an important problem that impacts the motion analysis approach developed under this doctoral research - the aperture problem.

When an object moves in a scene, it generates a changing intensity pattern across the retina, or image plane, of the viewer. If the object moves with constant velocity components (u, v) in a time interval δt , the intensity of a single point in the image plane can be represented by the function $f(x + \delta x, y + \delta y, t + \delta t)$ where $\delta x = u\delta t$, $\delta y = v\delta t$ and the velocity vector (u, v) represents the velocity of the object at the point x, y . Expanding this function in a Taylor series then yields

$$f(x + \delta x, y + \delta y, t + \delta t) = f(x, y, t) + \frac{\partial f}{\partial x} \delta x + \frac{\partial f}{\partial y} \delta y + \frac{\partial f}{\partial t} \delta t \quad (28)$$

where the higher order terms have been ignored and the partial derivatives are evaluated at (x, y, t) . The key assumption made in the optical flow derivation is that the intensity of a point $x + \delta x, y + \delta y$ at time $t + \delta t$ is identical to the intensity at x, y, t . That is, the moving pattern (or point in this case) simply shifts position in time. This assumption implies

$$f(x + \delta x, y + \delta y, t + \delta t) = f(x, y, t) \quad (29)$$

Cancelling terms in Equation 28, dividing by δt , and letting δt go to zero then gives,

$$-\frac{\partial f}{\partial t} = \frac{\partial f}{\partial x} \frac{dx}{dt} + \frac{\partial f}{\partial y} \frac{dy}{dt} \quad (30)$$

where the velocity of the moving point is specified by the terms $\frac{dx}{dt}$ and $\frac{dy}{dt}$. Letting $u = \frac{dx}{dt}$ and $v = \frac{dy}{dt}$, Equation 30 can be written as

$$-\frac{\partial f}{\partial t} = \frac{\partial f}{\partial x} u + \frac{\partial f}{\partial y} v \quad (31)$$

Or, representing the velocity components by the vector $\mathbf{u}$,

$$-\frac{\partial f}{\partial t} = \nabla f \cdot \mathbf{u} \quad (32)$$

where ∇f is the spatial gradient of the image. Equation 31 has an interesting interpretation. It implies that if a viewing point in the image plane is held fixed, the time rate of change in intensity of the image

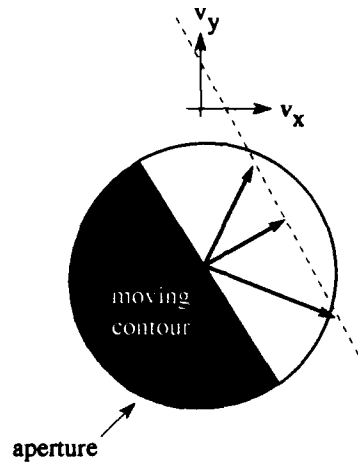


Figure 9. Direction ambiguity induced by a moving edge as seen through a circular aperture. The motion computed by Equation 32 is consistent with any of the shown velocity vectors (47).

point equals the spatial rate of change at the point multiplied by the velocity with which points in the scene move past the fixed viewing point.

The primary problem with the optical flow model as expressed in Equation 32, is that it is underconstrained (i.e., it yields only one component of motion in the direction of maximally increasing intensity). This shortcoming in the spatio-temporal gradient approach is often referred to as the "aperture problem" (47). For example, consider the moving brightness contour (edge) in Figure 9. Equation 32 gives the magnitude of the velocity component in the direction normal to the moving edge (i.e., the direction of ∇f); but, it does not provide any information about the magnitude of the velocity component lying parallel to the edge. Therefore, the actual direction of the movement is ambiguous and could lie along any of the vectors shown in the diagram. This same problem exists for a human observing a moving edge through an aperture. When the edge moves in a direction parallel to its boundary, it appears stationary to the observer.

The underconstrained spatio-temporal gradient flow model represented by Equation 32 is commonly solved by first imposing one or more of the motion constraints described earlier, and then solving for the velocity at each point in the scene using various combinatorial optimization techniques (4, 26, 27). Perhaps the most commonly used method imposes a non-linear, spatial smoothness constraint on the velocity field. This constraint, which requires that neighboring points in the image plane

have similar velocities (similar to the “physics” constraint introduced in the previous section), attempts to minimize the square of the magnitude of the gradient of the optical flow velocity:

$$\left(\frac{\partial u}{\partial x}\right)^2 + \left(\frac{\partial u}{\partial y}\right)^2 \quad \text{and} \quad \left(\frac{\partial v}{\partial x}\right)^2 + \left(\frac{\partial v}{\partial y}\right)^2 \quad (33)$$

In reality, however, neighboring points do not necessarily have similar velocities (such as at an object boundary or at an occluding edge); thus, in these instances the smoothness constraint leads to inaccurate velocity estimates. E. Hildreth (26) attempted to solve this problem by computing the velocity along boundary contours; however, her approach imposes a velocity smoothness constraint along the arc of the boundary that also leads to inaccurate motion estimates (4). Other problems commonly associated with the spatio-temporal gradient approach as implemented by Horn *et. al.* include 1) it does not allow discontinuities in the velocity field that occur when a moving object is temporarily occluded, and 2) imposing a smoothness constraint requires the computation of second order spatial derivatives which magnify noise in the scene. The motion estimation approach developed in this research performs better under these conditions by combining the flow “averaging” effects of the wavelet transform with the cooperative-competitive flow correction properties of a gated dipole filter (see Chapter V).

2.3.3 Fourier Phase Approach. The Fourier phase approach computes the optical flow of an object moving across a 2D scene by measuring the phase change associated with a purely translational shift in x and y over time. The approach takes advantage of the shift property of the Fourier transform which states that a shift in the spatial domain corresponds to a linear phase term in frequency domain. Or, more formally, (33):

$$\text{Fourier Shift Property: If } F(u, v) = \mathcal{F}\{f(x, y)\} \text{ and } g(x, y) = f(x - a, y - b), \text{ then} \\ G(u, v) = \mathcal{F}\{f(x - a, y - b)\} = F(u, v)e^{-i2\pi(ua+vb)}.$$

The Fourier phase approach generally works in the following way. Let $f(m, n)$ represent a discretely sampled signal traveling with a constant velocity. If the amplitude and phase of the DFT of $f(m, n)$ are given by $A_f(k, l)$ and $\phi_f(k, l)$ respectively, and if f translates a discrete distance $\Delta m, \Delta n$ in time Δt across an $M \times M$ image plane, then the above shift property implies the phase difference over the

discrete time interval can be expressed by

$$\begin{aligned}
 \Delta\phi(k, l) &= \phi_f(k, l) - \phi_g(k, l) \\
 &= \phi_f(k, l) - (\phi_f(k, l) - (\frac{2\pi k}{M}\Delta m + \frac{2\pi l}{M}\Delta n)) \\
 &= \frac{2\pi k}{M}\Delta m + \frac{2\pi l}{M}\Delta n
 \end{aligned} \tag{34}$$

If the phase change is known at two frequencies, Equation 34 can be solved for Δm and Δn which can then be divided by Δt to yield the velocity of the object. Equation 34 holds for a single object translating across an image plane (31). If multiple objects are present, the standard technique is to subdivide the scene into non-overlapping blocks and compute the phase shift within each block. Additionally, if noise is present in the scene, Equation 34 can be evaluated at several different frequencies (using a least-mean-squares method, for example) to find the shift parameters that best fit the phase data (23).

The primary problem with the Fourier phase approach lies in the ability to accurately estimate the phase shift in Equation 34. If the object displacement over the time interval Δt is large, the change in phase may be larger than 2π . Because the DFT wraps phase between $-\pi$ and π , discrete phase measurements of a fast object may yield a gross underestimate of the actual velocity. One approach to this problem is to evaluate Equation 34 at low frequencies where the longer spatial frequency periods allow larger displacements over time before the phase periodically repeats. This approach, however, prevents the accurate estimate of localized motion parameters. The optical flow algorithm in Chapter V overcomes this problem through the use of spatio-temporal wavelets which are localized in space and time. Another approach is to employ a phase unwrapping routine to unwrap the true phase difference from the measured, periodic phase components. Unfortunately, these routines rely on user-defined thresholds which make them difficult to generally apply to natural image sequences. For a good example of a phase unwrapping routine see Oppenheim and Schaffer (46).

2.4 Motion Analysis Using a Spatio-Temporal Frequency Approach

Experimental studies of motion analysis in the human visual system imply motion information is processed by the brain in parallel channels, each of which is selective for a specific location, spatial frequency, and velocity. Several researchers have attempted to model this behavior using a spatio-temporal Fourier analysis technique which decomposes a moving object into patches of oriented

sinusoids moving past an image plane region at some temporal frequency (15, 25, 57). This section reviews several key concepts associated with the computation of optical flow using spatio-temporal Fourier frequency measurements, beginning with a discussion of the connection between spatio-temporal frequency (STF) and the velocity of a 2D object moving across an image plane. It is shown that an STF component can be interpreted as a 2D sinusoid with a fixed spatial period and orientation moving with a constant velocity proportional to its temporal frequency. Furthermore, the Fourier transform of an object moving with constant speed and direction can be represented by a plane in Fourier frequency space. Finally, a relevant frequency filtering technique is described which uses STF information to compute optical flow.

Consider the single spatio-temporal frequency component shown in Figure 10a). If the frequency component is expressed as the delta pair $\delta(f_x - a, f_y - b, f_t - c) + \delta(f_x + a, f_y + b, f_t + c)$, then its inverse Fourier transform is given by the traveling wave:

$$f(x, y, t) = \cos(ax + by + ct) \quad (35)$$

where the amplitude of the inverse transform has been ignored. If the time-dependent cosine is evaluated at $t = 0$, then, following Goodman (21), it can be represented by a family of parallel lines of constant phase as shown in Figure 10b). Here, each line represents a locus of points along which $\cos(ax + by) = 1$ or, equivalently, $ax + by = 2k\pi$, $k = 0, 1, 2, \dots$. The velocity of the wave is depicted by the vector, $\mathbf{V}$, drawn perpendicular to the lines of constant phase in Figure 10b). Assuming the x and y velocity components of $\mathbf{V}$ are given by v_x and v_y , the traveling cosine wave can be expressed as

$$f(x, y, t) = \cos(a(x - v_x t) + b(y - v_y t)) \quad (36)$$

where c has been defined by $c = -av_x - bv_y$. Combining the ratio $\frac{v_y}{v_x} = \frac{b}{a}$ obtained from like triangles in Figure 10b), with Equations 35 and 36, yields the following relationship between the velocity components of the wave and the spatio-temporal frequencies (a, b, c)

$$\begin{aligned} v_x &= -\frac{ac}{a^2 + b^2} \\ v_y &= -\frac{bc}{a^2 + b^2} \end{aligned} \quad (37)$$

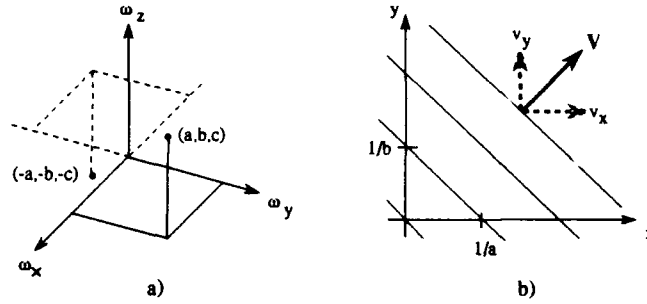


Figure 10. a) Delta pair in spatio-temporal frequency space. b) Lines of constant phase associated with inverse Fourier transform of delta pair in part a).

Thus, the spatial frequency pair (a, b) define the pitch, orientation and direction of the traveling wave, while its speed is directly proportional to the temporal frequency c (19). This implies, a 3D sequence of images can be constructed from the summation of appropriately shifted and scaled 2D sines and cosines moving at different speeds. Now consider the behavior in spatio-temporal frequency space of an object, rather than a single spatial frequency, traveling across a 2D image plane.

Assume a stationary object is imaged onto a viewing plane and that the intensity of the object at some point in the image is described by the function $f(x, y)$ (i.e., its intensity does not vary with time). Next, assume the object translates along a linear trajectory with constant velocity (v_x, v_y) . The motion of the intensity pattern as it sweeps across the image plane can then be modeled by the function $f(x - v_x t, y - v_y t)$ (58). Now consider the spatio-temporal Fourier transform of this function.

$$\mathcal{F}\{f(x - v_x t, y - v_y t)\} = \int \int \int_{-\infty}^{\infty} f(x - v_x t, y - v_y t) e^{-i2\pi(f_x x + f_y y + f_t t)} dx dy dt \quad (38)$$

where f_x, f_y, f_t are spatial and temporal frequencies. Using the substitution variables

$$p = x - v_x t$$

$$q = y - v_y t$$

the Fourier transform of the moving object can be expressed as

$$\mathcal{F}\{f(x - v_x t, y - v_y t)\} = \int \int \int_{-\infty}^{\infty} f(p, q) e^{-i2\pi[f_x(p + v_x t) + f_y(q + v_y t) + f_t t]} dp dq dt$$

$$\begin{aligned}
&= \int \int \int_{-\infty}^{\infty} f(p, q) e^{-i2\pi[f_x p + f_y q + (f_t + v_x f_x + v_y f_y)t]} dp dq dt \\
&= \int_{-\infty}^{\infty} F(f_x, f_y) e^{-i2\pi(f_t + v_x f_x + v_y f_y)t} dt \\
&= F(f_x, f_y) \cdot \delta(f_t + v_x f_x + v_y f_y)
\end{aligned} \tag{39}$$

where $F(f_x, f_y)$ is the Fourier transform of the stationary object $f(x, y)$. Equation 39 implies that when an object moves in space with a constant velocity, each spatial frequency component of its static Fourier transform (i.e., its Fourier transform when stationary in space) simply shifts along the temporal frequency axis by the amount $v_x f_x + v_y f_y$. To help visualize this behavior, consider the one dimensional function shown in Figure 11.

Figure 11 a) shows a stationary rectangle function, $\text{rect}(x, t)$, in an image plane coordinate system. Since the function is stationary and the shape of the rect does not change in time, its spatio-temporal Fourier transform (Figure 11 b) is restricted to the f_x axis. In Figure 11 c), the rectangle moves with some constant velocity v_x . The Fourier transform of the moving rect, shown in Figure 11 d), is then given by $F(f_x) \cdot \delta(f_t + v_x f_x)$. Thus, constant velocity motion in one spatial dimension shifts the Fourier transform of the stationary object to a *line* in 2D spatio-temporal frequency space defined by $f_t = -v_x f_x$. Similarly, constant velocity motion in *two* dimensions shifts the Fourier transform of the object onto a *plane* in 3D frequency space defined by $f_t = -(v_x f_x + v_y f_y)$ (see Figure 12).

A single temporal frequency component in the “velocity” spectrum of the moving object can also be expressed as

$$\begin{aligned}
f_t &= -(v_x f_x + v_y f_y) \\
&= -\mathbf{v} \cdot \mathbf{f} \\
&= -v f \cos(\theta - \alpha)
\end{aligned} \tag{40}$$

where v and θ are the speed and direction of motion respectively, and f and α are the magnitude and direction of the spatial frequency f_x, f_y . Equation 40 specifies the magnitude of the velocity component lying in the direction of the spatial frequency. As one equation in two unknowns (v and θ), it cannot unambiguously identify the velocity of the moving object. That is, like the spatio-temporal gradient approach, this method also suffers from the “aperture” problem.

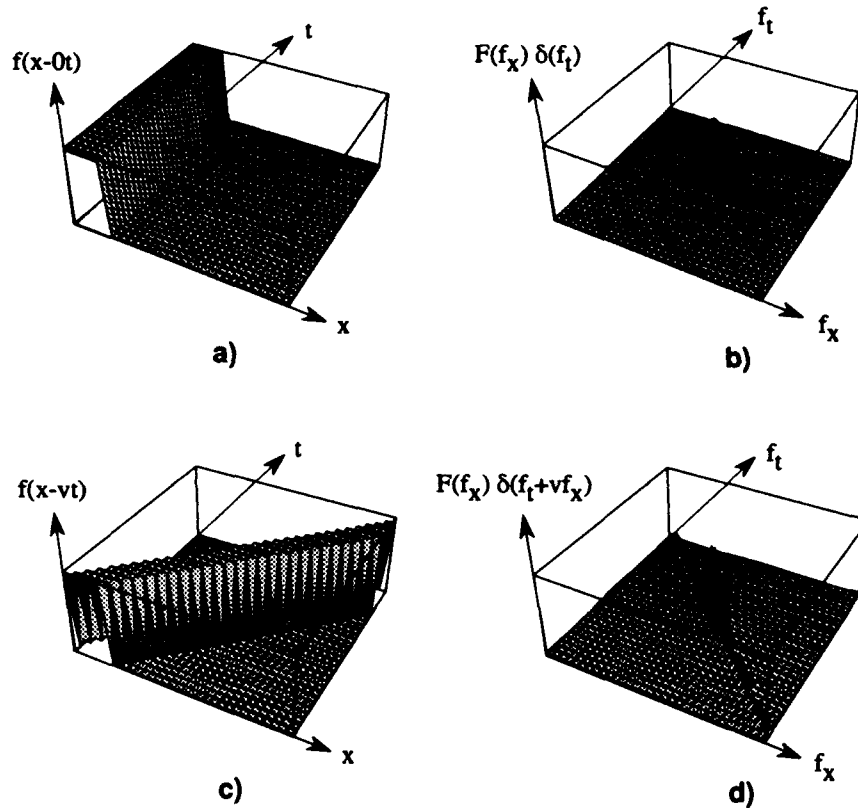


Figure 11. a) and b) Stationary one dimensional rectangle function and its spatio-temporal Fourier transform. c) and d) Rectangle function moving with constant velocity, v , and its Fourier transform.

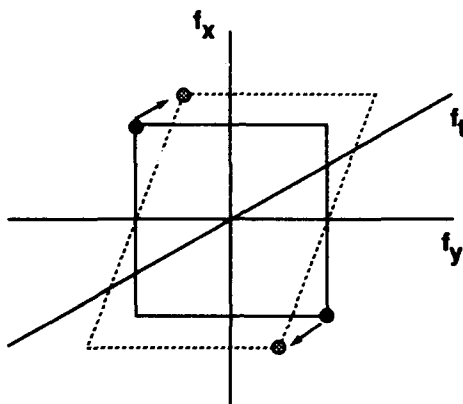


Figure 12. Dashed lines outline plane generated in 3D spatio-temporal frequency space by a two dimensional object moving with constant vertical velocity. Circles illustrate shift of a single frequency component.

One possible way to solve this problem is to measure the temporal frequencies f_t^1, f_t^2 associated with two different spatial frequency pairs f_x^1, f_y^1 and f_x^2, f_y^2 . This generates two equations which can be used to simultaneously solve for the two unknowns v_x and v_y . In previous efforts, researchers have filtered the time varying image with a bank of spatially localized spatio-temporal filters (e.g., Gabor filters), where each filter is tuned to a different spatial and temporal frequency as shown in Figure 13 (25, 58). As an image moves past a receptive field, the spatio-temporal filters that match the texture content and speed of the image at a given location will activate, yielding several spatial and temporal frequency triplets which can then be used to determine the velocity vector associated with that receptive field.

The major advantage of the Fourier motion analysis method is that it solves the aperture problem without imposing the artificial constraints used in the spatio-temporal gradient optimization approach. This has the potential of making the technique less susceptible to noise, and more accurate at velocity boundaries such as occlusions or object edges in the scene. Additionally, the Fourier technique more closely models the early stages of the human visual system by creating individual motion detection channels that select for a given spatial frequency, direction and speed. The major disadvantages associated with existing STF approaches are 1) they rely on "heuristic" filter banks that provide little control over important inter-dependent filter design characteristics such as filter overlap, filter

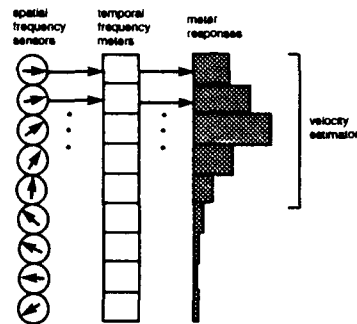


Figure 13. Bank of spatio-temporal Gabor filters used by Watson and Ahumada to compute local velocity components (58).

bandwidth, and space/spatial frequency localization, 2) they employ short-time Fourier transform techniques that limit their analysis to a fixed resolution in space and time, 3) their filter designs cannot be easily modified to match the design constraints imposed by different problem scenarios, and 4) their flow computation algorithms have not been demonstrated in the presence of noise. The spatio-temporal frequency approach developed in the following chapters overcomes these problems through the use of a rigorous, wavelet-based mathematical framework and a cooperative-competitive flow restoration methodology.

2.5 Conclusion

Conventional methods for characterizing motion in 3D imagery are susceptible to noise and require the use of motion constraints that reduce their accuracy at boundaries where the velocity field is discontinuous. Spatio-temporal frequency techniques overcome many of these problems by computing optical flow from multiples image frames. Current spatio-temporal frequency techniques employ “heuristic” filter banks that are not easily adaptable to natural imagery. Additionally, these methods are based on the Short Time Fourier Transform, which restricts the analysis to a single scale in space and time. Finally, both the spatio-temporal gradient and spatio-temporal frequency motion analysis methods are computationally expensive. The wavelet multiresolution analysis yields a mathematically rigorous method for decomposing an image into a sequence of approximation and detail spaces that capture unique object characteristics at multiple spatial scales. It is implemented as

a fast, sub-band coding scheme which generates a non-uniform filter bank of independent, *spatially* oriented frequency analysis channels. In the following chapters, the 2D multiresolution analysis will be extended to three dimensions to yield an efficient algorithm for decomposing 3D imagery into a set of *spatio-temporally* oriented frequency channels. The conventional 3D multiresolution analysis will be modified to enhance the flexibility of the separable filter design process and to increase the motion selectivity of the analyzing wavelets. Unlike existing motion analysis algorithms, the modified 3D wavelet decomposition algorithm allows one to rapidly compute optical flow over multiple scales in space and time, thereby creating a powerful tool for extracting moving targets in a scene based on their size, speed and direction.

III. A Wavelet Multiresolution Analysis for $L_2(\mathbb{R}^3)$

3.1 Introduction

Although Y. Meyer developed the general theory for wavelet multiresolution analyses in $L_2(\mathbb{R}^n)$, his work does not provide details for actually constructing orthonormal wavelet bases for these spaces. Additionally, previous instantiations of Meyer's wavelet multiresolution analysis dealt exclusively with one and two-dimensional signals (14, 39). Thus, this chapter begins by extending Mallat's theorems for the construction of wavelet orthonormal bases for $L_2(\mathbb{R})$ and $L_2(\mathbb{R}^2)$ to the space of finite energy spatio-temporal signals, $L_2(\mathbb{R}^3)$. It is shown that a separable wavelet orthonormal basis for $L_2(\mathbb{R}^3)$ consists of a set of seven dyadic wavelets evaluated over all possible integer shifts and dilations. The second section of the chapter presents an "oct-tree" sub-band coding scheme for implementing the Discrete Spatio-temporal Wavelet Transform. The algorithm generates a bank of octave-band filters such that each filter possesses uniform spatial and temporal frequency characteristics. The sub-band decomposition algorithm is then applied to a set of synthetic 3D imagery to demonstrate its ability to extract vertical, horizontal or diagonal features from moving or stationary targets. The chapter concludes with a discussion of the advantages and disadvantages of using the "conventional" wavelet multiresolution analysis for 3D motion analysis.

3.2 Orthonormal Wavelet Basis

The vector space $L_2(\mathbb{R}^3)$ consists of all functions, $f(x, y, t)$ such that

$$\int \int \int_{-\infty}^{\infty} |f(x, y, t)|^2 dx dy dt < \infty \quad (41)$$

A wavelet multiresolution analysis for $L_2(\mathbb{R}^3)$ consists of a chain of approximation spaces, V_j , that differ in resolution by a factor of 2^j in each of the three dimensions x , y , and t . These spaces satisfy an extension of the $L_2(\mathbb{R})$ multiresolution analysis properties listed in Section 2.2.3. In particular, their union is dense in $L_2(\mathbb{R}^3)$ and their intersection contains only the zero element. Like the multiresolution analysis for $L_2(\mathbb{R})$, the approximation of a three-dimensional signal at the j th resolution level is obtained by orthogonally projecting the signal onto V_j . The details between the j th and $j + 1$ st

approximations are captured in the detail space W_j where

$$V_{j+1} = V_j \oplus W_j \quad (42)$$

and $\oplus$ denotes the direct sum operation. Each of the approximation spaces contains a three-dimensional scaling function $\phi(2^j x, 2^j y, 2^j t)$ where the set $\{\phi(2^j x - l, 2^j y - m, 2^j t - n) \mid (l, m, n) \in \mathbb{Z}^3\}$ forms an orthonormal basis for V_j .

V_j is a closed, linear subspace of $L_2(\mathbb{R}^3)$, and is formed from the tensor product of three identical $L_2(\mathbb{R})$ approximation spaces (42). The term “identical” here implies each space is constructed from the same scaling function. If V_x^j, V_y^j and V_t^j are identical approximation spaces in $L_2(\mathbb{R})$, then the j th approximation space in $L_2(\mathbb{R}^3)$ is defined by

$$V_j = V_x^j \otimes V_y^j \otimes V_t^j = \overline{\text{Span}\{F(x, y, t) = f(x)g(y)h(t) \mid f \in V_x^j, g \in V_y^j \text{ and } h \in V_t^j\}} \quad (43)$$

where $\otimes$ denotes the tensor product operation. The unique scaling function for V_j is given by the separable product

$$\Phi(x, y, t) = \phi(x)\phi(y)\phi(t) \quad (44)$$

where $\phi(x), \phi(y)$ and $\phi(t)$ are identical scaling functions in $L_2(\mathbb{R})$. It is not difficult to show that the set $\{2^{\frac{3j}{2}}\phi(2^j x - l)\phi(2^j y - m)\phi(2^j t - n) \mid (l, m, n) \in \mathbb{Z}^3\}$ then forms an orthonormal basis for V_j . Furthermore, if $\psi(x), \psi(y)$ and $\psi(t)$ represent the wavelets generated by the $L_2(\mathbb{R})$ scaling functions $\phi(x), \phi(y)$ and $\phi(t)$, then the following theorem shows one can construct dyadic wavelet orthonormal bases for W_j and $L_2(\mathbb{R}^3)$ from seven sets of scaled and shifted “wavelets.”

Theorem 1. Let ψ be the one-dimensional wavelet generated by the scaling function ϕ . Then the seven “wavelets”

$$\begin{aligned} \Psi_j^1(x, y, t) &= 2^{\frac{3j}{2}}\phi(2^j x)\phi(2^j y)\psi(2^j t) \\ \Psi_j^2(x, y, t) &= 2^{\frac{3j}{2}}\phi(2^j x)\psi(2^j y)\phi(2^j t) \\ \Psi_j^3(x, y, t) &= 2^{\frac{3j}{2}}\phi(2^j x)\psi(2^j y)\psi(2^j t) \\ \Psi_j^4(x, y, t) &= 2^{\frac{3j}{2}}\psi(2^j x)\phi(2^j y)\phi(2^j t) \\ \Psi_j^5(x, y, t) &= 2^{\frac{3j}{2}}\psi(2^j x)\phi(2^j y)\psi(2^j t) \end{aligned}$$

$$\begin{aligned}
\Psi_j^6(x, y, t) &= 2^{\frac{3j}{2}} \psi(2^j x) \psi(2^j y) \phi(2^j t) \\
\Psi_j^7(x, y, t) &= 2^{\frac{3j}{2}} \psi(2^j x) \psi(2^j y) \psi(2^j t)
\end{aligned} \tag{45}$$

are such that for each $j \in \mathbb{Z}$, $\{\Psi_j^p(x-l, y-m, t-n) \mid (l, m, n) \in \mathbb{Z}^3; p = 1, 2, \dots, 7\}$ forms an orthonormal basis for W_j and the set $\{\Psi_j^p(x-l, y-m, t-n) \mid j \in \mathbb{Z}; (l, m, n) \in \mathbb{Z}^3; p = 1, 2, \dots, 7\}$ forms an orthonormal basis for $L_2(\mathbb{R}^3)$.

Proof. Let $V_j, j \in \mathbb{Z}$, be a multiresolution approximation of $L_2(\mathbb{R}^3)$ formed by the tensor product

$$V_j = V_j^x \otimes V_j^y \otimes V_j^t = \overline{\text{Span}\{F(x, y, t) = f(x)g(y)h(t) \mid f \in V_j^x, g \in V_j^y \text{ and } h \in V_j^t\}} \tag{46}$$

where V_j^x, V_j^y and V_j^t are multiresolution approximations of $L_2(\mathbb{R})$. Let W_j^x, W_j^y and W_j^t be the orthogonal complements of the closed, linear spaces $V_j^x \subset V_{j+1}^x, V_j^y \subset V_{j+1}^y$ and $V_j^t \subset V_{j+1}^t$. Then

$$\begin{aligned}
V_{j+1} &= V_{j+1}^x \otimes V_{j+1}^y \otimes V_{j+1}^t \\
&= (W_j^x \oplus V_j^x) \otimes (W_j^y \oplus V_j^y) \otimes (W_j^t \oplus V_j^t)
\end{aligned} \tag{47}$$

The right hand side of Equation 47 can be rewritten as follows

$$\begin{aligned}
\text{RHS} &= [W_j^x \otimes W_j^y \otimes W_j^t] \oplus [W_j^x \otimes W_j^y \otimes V_j^t] \oplus [W_j^x \otimes V_j^y \otimes W_j^t] \\
&\oplus [W_j^x \otimes V_j^y \otimes V_j^t] \oplus [V_j^x \otimes W_j^y \otimes W_j^t] \oplus [V_j^x \otimes W_j^y \otimes V_j^t] \\
&\oplus [V_j^x \otimes V_j^y \otimes W_j^t] \oplus [V_j^x \otimes V_j^y \otimes V_j^t]
\end{aligned} \tag{48}$$

Since $V_j = V_j^x \otimes V_j^y \otimes V_j^t$, the orthogonal complement, W_j , of V_j in V_{j+1} can be expressed as

$$\begin{aligned}
W_j = V_{j+1} - V_j &= [W_j^x \otimes W_j^y \otimes W_j^t] \oplus [W_j^x \otimes W_j^y \otimes V_j^t] \oplus [W_j^x \otimes V_j^y \otimes W_j^t] \\
&\oplus [W_j^x \otimes V_j^y \otimes V_j^t] \oplus [V_j^x \otimes W_j^y \otimes W_j^t] \oplus [V_j^x \otimes W_j^y \otimes V_j^t] \\
&\oplus [V_j^x \otimes V_j^y \otimes W_j^t]
\end{aligned} \tag{49}$$

The sets of functions $\{2^{\frac{j}{2}} \phi(2^j x - l) \mid l \in \mathbb{Z}\}$, $\{2^{\frac{j}{2}} \phi(2^j y - m) \mid m \in \mathbb{Z}\}$, and $\{2^{\frac{j}{2}} \phi(2^j t - n) \mid n \in \mathbb{Z}\}$, form orthonormal bases respectively for the $L_2(\mathbb{R})$ approximation spaces V_j^x, V_j^y and V_j^t . Additionally,

the sets of functions $\{2^{\frac{j}{2}}\psi(2^j x - l) \mid l \in \mathbb{Z}\}$, $\{2^{\frac{j}{2}}\psi(2^j y - m) \mid m \in \mathbb{Z}\}$, and $\{2^{\frac{j}{2}}\psi(2^j t - n) \mid n \in \mathbb{Z}\}$, form orthonormal bases respectively for the complementary spaces W_j^x , W_j^y and W_j^t . Thus, the set of functions $\{\Psi_j^p(x - l, y - m, t - n) \mid (l, m, n) \in \mathbb{Z}^3; p = 1, 2, \dots, 7\}$ forms an orthonormal bases for W_j . Furthermore, the fact that $L_2(\mathbb{R}^3)$ can be formed by the direct sum decomposition

$$\overline{\bigoplus_{j \in \mathbb{Z}} W_j} = L_2(\mathbb{R}^3) \quad (50)$$

implies the family of functions $\{\Psi_j^p(x - l, y - m, t - n) \mid j \in \mathbb{Z}; (l, m, n) \in \mathbb{Z}^3; p = 1, 2, \dots, 7\}$ constitutes an orthonormal basis for $L_2(\mathbb{R}^3)$. **Q.E.D.**

The approximation and detail signals at the j th resolution are obtained by orthogonally projecting the signal onto either V_j or W_j . Now, consider the projection of the signal onto V_j . If $\phi_{j;(l,m,n)}(x, y, t) = 2^{\frac{3j}{2}}\phi(2^j x - l)\phi(2^j y - m)\phi(2^j t - n)$, then the orthogonal projection, $A_j f$, of the signal f onto the approximation space V_j can be represented by the series

$$A_j f = \sum_l \sum_m \sum_n a_{j;(l,m,n)} \phi_{j;(l,m,n)} \quad (51)$$

where the projection coefficient $a_{j;(l,m,n)}$ is given by the inner product of the signal with the orthonormal basis element $2^{\frac{3j}{2}}\phi(2^j x - l)\phi(2^j y - m)\phi(2^j t - n)$, i.e.,

$$a_{j;(l,m,n)} = \int \int \int_{-\infty}^{\infty} f(x, y, t) 2^{\frac{3j}{2}}\phi(2^j x - l)\phi(2^j y - m)\phi(2^j t - n) dx dy dt \quad (52)$$

The orthogonal projection of the signal onto the j th detail space, W_j , is obtained in a similar manner; however, the signal must now be projected onto each of the seven basis subsets described in Theorem 1. Thus, if $\psi_{j;(l,m,n)}^p(x, y, t) = \psi_j^p(x - l, y - m, t - n)$, then, by Theorem 1, the detail signal $D_j f$ can be expressed by the series

$$D_j f = \sum_l \sum_m \sum_n \sum_p d_{j;(l,m,n)}^p \psi_{j;(l,m,n)}^p \quad (53)$$

where $p = 1, 2, \dots, 7$ and the detail coefficients $d_{j;(l,m,n)}^p$ are given by the seven inner products

$$d_{j;(l,m,n)}^1 = \int \int \int_{-\infty}^{\infty} f(x, y, t) 2^{\frac{3j}{2}}\phi(2^j x - l)\phi(2^j y - m)\psi(2^j t - n) dx dy dt$$

$$\begin{aligned}
d_{j;(l,m,n)}^2 &= \int \int \int_{-\infty}^{\infty} f(x,y,t) 2^{\frac{3j}{2}} \phi(2^j x - l) \psi(2^j y - m) \phi(2^j t - n) dx dy dt \\
d_{j;(l,m,n)}^3 &= \int \int \int_{-\infty}^{\infty} f(x,y,t) 2^{\frac{3j}{2}} \phi(2^j x - l) \psi(2^j y - m) \psi(2^j t - n) dx dy dt \\
d_{j;(l,m,n)}^4 &= \int \int \int_{-\infty}^{\infty} f(x,y,t) 2^{\frac{3j}{2}} \psi(2^j x - l) \phi(2^j y - m) \phi(2^j t - n) dx dy dt \\
d_{j;(l,m,n)}^5 &= \int \int \int_{-\infty}^{\infty} f(x,y,t) 2^{\frac{3j}{2}} \psi(2^j x - l) \phi(2^j y - m) \psi(2^j t - n) dx dy dt \\
d_{j;(l,m,n)}^6 &= \int \int \int_{-\infty}^{\infty} f(x,y,t) 2^{\frac{3j}{2}} \psi(2^j x - l) \psi(2^j y - m) \phi(2^j t - n) dx dy dt \\
d_{j;(l,m,n)}^7 &= \int \int \int_{-\infty}^{\infty} f(x,y,t) 2^{\frac{3j}{2}} \psi(2^j x - l) \psi(2^j y - m) \psi(2^j t - n) dx dy dt \quad (54)
\end{aligned}$$

Since the sets of shifted scaling functions and wavelets form orthonormal bases for the approximation and detail spaces V_j and W_j respectively, the signal projections onto either of these L_2 spaces are uniquely and completely represented by the l_2 coefficients in Equations 52 and 54. The next section describes a pyramidal algorithm that enables one to quickly and efficiently compute these coefficients for multiple spatial and temporal scales from a sampled version of the input signal.

3.3 Discrete Multiresolution Decomposition Algorithm

This section describes a fine-to-coarse digital algorithm for computing the approximation and detail coefficients associated with an $L_2(\mathbb{R}^3)$ wavelet multiresolution analysis. The algorithm is constructed through an extension of Mallat's 1D decomposition algorithm (40). It begins by assuming the sequence obtained by sampling the signal in x, y and t represents the coefficients associated with the orthogonal projection of f onto the "zeroth" approximation space V_0 . The coefficients of the next lower resolution approximation and detail signals, $A_{-1}f$ and $D_{-1}f$, are computed first by convolving the discrete input signal with the 3D separable QMF pair derived below and keeping every other sample in x, y and t (i.e., decimating the output by a factor of two). This process is repeatedly applied to the approximation coefficients generated at each resolution level to obtain the detail and approximation coefficients at successively lower levels.

Since the chain of approximation spaces $\{V_j \mid j \in \mathbb{Z}\}$ forms a multiresolution analysis in $L_2(\mathbb{R}^3)$, the first property in Section 2.2.3 guarantees that $V_j \subset V_{j+1}$ for all j . This implies that for any triplet $(l, m, n) \in \mathbb{Z}^3$, the basis element $2^{\frac{3j}{2}} \phi(2^j x - l) \phi(2^j y - m) \phi(2^j t - n)$ in V_j can be

expanded in the orthonormal basis of V_{j+1} . Therefore, one can write

$$\begin{aligned} \phi(2^j x - l)\phi(2^j y - m)\phi(2^j t - n) &= 2^{3(j+1)} \sum_p \sum_q \sum_r \langle \phi(2^j u - l)\phi(2^j v - m)\phi(2^j w - n), \\ &\phi(2^{j+1} u - p)\phi(2^{j+1} v - q)\phi(2^{j+1} w - r) \rangle \phi(2^{j+1} x - p)\phi(2^{j+1} y - q)\phi(2^{j+1} t - r) \end{aligned} \quad (55)$$

where $p, q, r \in \mathbb{Z}$. Expanding the inner product in the above expression in its integral form then gives

$$\begin{aligned} 2^{3(j+1)} \langle \cdot, \cdot \rangle &= 2^{3(j+1)} \int \int \int_{-\infty}^{\infty} [\phi(2^j u - l)\phi(2^j v - m)\phi(2^j w - n)] \\ &\cdot [\phi(2^{j+1} u - p)\phi(2^{j+1} v - q)\phi(2^{j+1} w - r)] du dv dw \end{aligned} \quad (56)$$

Using the variable substitutions

$$\begin{aligned} \frac{a}{2} &= 2^j u - l \\ \frac{b}{2} &= 2^j v - m \\ \frac{c}{2} &= 2^j w - n \end{aligned} \quad (57)$$

the right hand side of Equation 56 can be rewritten as

$$\int \int \int_{-\infty}^{\infty} [\phi(\frac{a}{2})\phi(\frac{b}{2})\phi(\frac{c}{2})][\phi(a - (p - 2l))\phi(b - (q - 2m))\phi(c - (r - 2n))] da db dc \quad (58)$$

Next, defining the function h by

$$h(x) \equiv \int_{-\infty}^{\infty} \phi(\frac{y}{2})\phi(y - x) dy \quad (59)$$

the right hand side of Equation 56 can be combined with Equation 59 to yield the following expression for the arbitrary basis element $2^{\frac{3j}{2}} \phi(2^j x - l)\phi(2^j y - m)\phi(2^j t - n)$

$$\begin{aligned} 2^{\frac{3j}{2}} \phi(2^j x - l)\phi(2^j y - m)\phi(2^j t - n) &= \\ \sum_p \sum_q \sum_r \bar{h}(2l - p)\bar{h}(2m - q)\bar{h}(2n - r) &\phi(2^{j+1} x - p)\phi(2^{j+1} y - q)\phi(2^{j+1} t - r) \end{aligned} \quad (60)$$

where $\bar{h}(x) = h(-x)$. In order to obtain the coefficients associated with the signal projection onto V_j , one can form the inner product of f with the arbitrary basis element in Equation 60 as follows

$$\begin{aligned}
a_{j;(l,m,n)} &= \langle f, 2^{\frac{3j}{2}} \phi(2^j x - l) \phi(2^j y - m) \phi(2^j t - n) \rangle \\
&= \sum_p \sum_q \sum_r \bar{h}(2l - p) \bar{h}(2m - q) \bar{h}(2n - r) \langle f, \phi(2^{j+1} x - p) \phi(2^{j+1} y - q) \phi(2^{j+1} t - r) \rangle \\
&= \sum_p \sum_q \sum_r \bar{h}(2l - p) \bar{h}(2m - q) \bar{h}(2n - r) a_{j+1;(p,q,r)} \\
&= [a_{j+1;(p,q,r)} * \bar{h}(p) * \bar{h}(q) * \bar{h}(r)] (2l, 2m, 2n) \quad (61)
\end{aligned}$$

where “*” is the discrete convolution operator. Equation 61 shows the discrete representation of the orthogonal projection of the signal onto the approximation space V_j can be obtained by discretely convolving the coefficients of the projection onto the next higher resolution level, V_{j+1} , with the separable impulse response $h(-p)h(-q)h(-r)$ and keeping every other sample in each dimension. Next, a similar procedure is used to derive the coefficients associated with the orthogonal projection onto the detail space, D_j^7 spanned by the functions $\{2^{\frac{3j}{2}} \psi(2^j x - l) \psi(2^j y - m) \psi(2^j t - n) \mid (l, m, n) \in \mathbb{Z}^3\}$. These results can then be generalized to compute the coefficients associated with the projections onto the remaining detail spaces $D_j^1 - D_j^6$.

Recall that the approximation space V_{j+1} is formed by the direct sum of V_j and W_j . Thus, W_j is contained in V_{j+1} and any basis element in W_j can be expanded in the orthonormal basis of V_{j+1} . In particular, the W_j basis element $\psi(2^j x - l) \psi(2^j y - m) \psi(2^j t - n)$ can be written as

$$\begin{aligned}
\psi(2^j x - l) \psi(2^j y - m) \psi(2^j t - n) &= 2^{3(j+1)} \sum_p \sum_q \sum_r \langle \psi(2^j u - l) \psi(2^j v - m) \psi(2^j w - n), \\
&\quad \phi(2^{j+1} u - p) \phi(2^{j+1} v - q) \phi(2^{j+1} w - r) \rangle \phi(2^{j+1} x - p) \phi(2^{j+1} y - q) \phi(2^{j+1} t - r) \quad (62)
\end{aligned}$$

Using the variable substitutions in Equation 57 and following the same procedure used earlier, the inner product in the above expression becomes

$$\int \int \int_{-\infty}^{\infty} [\phi(\frac{a}{2}) \phi(\frac{b}{2}) \phi(\frac{c}{2})] [\phi(a - (p - 2l)) \phi(b - (q - 2m)) \phi(c - (r - 2n))] da db dc \quad (63)$$

Defining the impulse response $g(x)$ by

$$g(x) \equiv \int_{-\infty}^{\infty} \psi\left(\frac{y}{2}\right) \psi(y-x) dy \quad (64)$$

and combining it with the integral in Equation 63 yields

$$2^{\frac{3j}{2}} \psi(2^j x - l) \psi(2^j y - m) \psi(2^j t - n) = \sum_p \sum_q \sum_r \bar{g}(2l - p) \bar{g}(2m - q) \bar{g}(2n - r) \phi(2^{j+1}x - p) \phi(2^{j+1}y - q) \phi(2^{j+1}t - r) \quad (65)$$

where $\bar{g}(x) = g(-x)$. The detail coefficients for the orthogonal projection onto D_j^7 are now obtained by taking the inner product of f with the W_j basis element $2^{\frac{3j}{2}} \psi(2^j x - l) \psi(2^j y - m) \psi(2^j t - n)$ as follows

$$\begin{aligned} d_{j;(l,m,n)}^7 &= \langle f, 2^{\frac{3j}{2}} \psi(2^j x - l) \psi(2^j y - m) \psi(2^j t - n) \rangle \\ &= \sum_p \sum_q \sum_r \bar{g}(2l - p) \bar{g}(2m - q) \bar{g}(2n - r) \langle f, \phi(2^{j+1}x - p) \phi(2^{j+1}y - q) \phi(2^{j+1}t - r) \rangle \\ &= \sum_p \sum_q \sum_r \bar{g}(2l - p) \bar{g}(2m - q) \bar{g}(2n - r) a_{j+1;(p,q,r)} \\ &= [a_{j+1;(p,q,r)} * \bar{g}(p) * \bar{g}(q) * \bar{g}(r)] (2l, 2m, 2n) \end{aligned} \quad (66)$$

Equation 66 shows the discrete representation of the orthogonal projection of the signal onto the portion of the detail space spanned by integer translations of the “wavelet” Ψ_j^7 can be obtained by discretely convolving the coefficients of the projection onto the next higher resolution approximation level, V_{j+1} , with the separable impulse response $g(-p)g(-q)g(-r)$ and keeping every other sample in each dimension. Following a similar procedure, the remaining detail coefficients $d_{j;(l,m,n)}^1$ through $d_{j;(l,m,n)}^6$ are given by the 3D discrete-space convolutions

$$\begin{aligned} d_{j;(l,m,n)}^1 &= [a_{j+1;(p,q,r)} * \bar{h}(p) * \bar{h}(q) * \bar{g}(r)] (2l, 2m, 2n) \\ d_{j;(l,m,n)}^2 &= [a_{j+1;(p,q,r)} * \bar{h}(p) * \bar{g}(q) * \bar{h}(r)] (2l, 2m, 2n) \\ d_{j;(l,m,n)}^3 &= [a_{j+1;(p,q,r)} * \bar{h}(p) * \bar{g}(q) * \bar{g}(r)] (2l, 2m, 2n) \\ d_{j;(l,m,n)}^4 &= [a_{j+1;(p,q,r)} * \bar{g}(p) * \bar{h}(q) * \bar{h}(r)] (2l, 2m, 2n) \\ d_{j;(l,m,n)}^5 &= [a_{j+1;(p,q,r)} * \bar{g}(p) * \bar{h}(q) * \bar{g}(r)] (2l, 2m, 2n) \end{aligned}$$

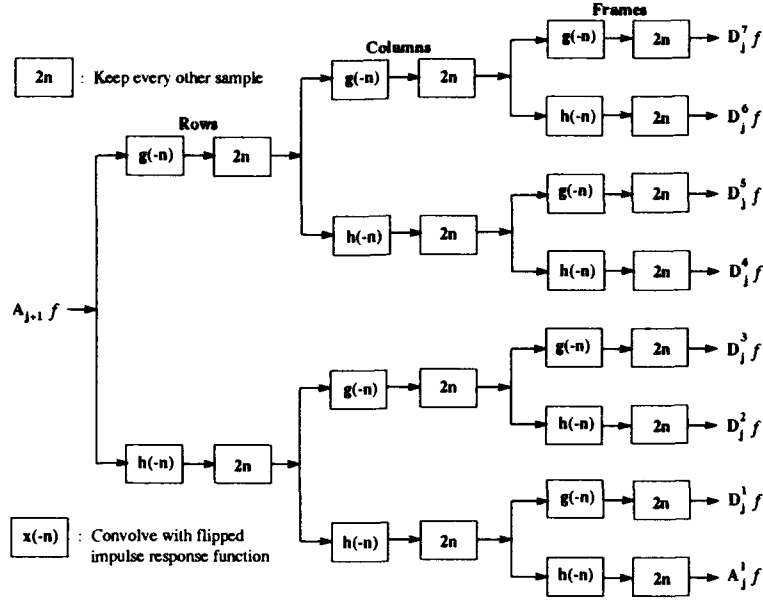


Figure 14. Oct-tree sub-band coding structure used to decompose the $j + 1$ st approximation coefficients into the j th approximation and detail coefficients in a conventional $L_2(\mathbb{R}^3)$ wavelet multiresolution analysis.

$$d_{j;(l,m,n)}^6 = [a_{j+1;(p,q,r)} * \bar{g}(p) * \bar{g}(q) * \bar{h}(r)](2l, 2m, 2n) \quad (67)$$

Equations 61, 66 and 67 show that the j th approximation and detail coefficients are obtained in a pyramidal fashion by discretely convolving the $j + 1$ st approximation coefficients with various combinations of the 1D impulse response pairs h and g and decimating the outputs by a factor of two. This process can be succinctly represented by the oct-tree sub-band coding structure shown in Figure 14. In a conventional extension of the $L_2(\mathbb{R}^2)$ wavelet multiresolution analysis, the oct-tree structure is formed by appending the canonical binary tree structure to each output of the quad-tree. The third tier convolves $h(-n)$ and $g(-n)$ with the frames of the $j + 1$ st approximation tensor where a frame represents a snapshot of the spatio-temporal signal at an instant in time. The next section describes the octave-band spatio-temporal frequency bank generated by this algorithm.

3.4 Spatio-Temporal Filter Bank Representation

The sequences $h(n)$ and $g(n)$ are the impulse responses of a QMF pair; thus, their z -transforms represent low and band-pass filters respectively (40). By repeatedly convolving a discrete 1D signal

with $h(-n)$ and $g(-n)$ and downsampling the outputs of each stage by a factor of two, the frequency content of the signal is effectively partitioned into octave-band regions of support. The binary tree decomposition algorithm developed by Mallat to generate the coefficients associated with orthogonal projections onto approximation and detail spaces can therefore be viewed as a sub-band filtering process in which the bandwidth and center frequency of each successive filter (moving out along the frequency axis) increases by a factor of two (56).

In the $L_2(\mathbb{R}^3)$ oct-tree coding structure, the impulse responses are convolved separately with each of the discrete spatial and temporal axes. By varying the order in which the impulse responses are applied to the rows, columns and frames of the signal, one can control the frequency characteristics of the corresponding sub-band filter. For example, convolving $h(n)$ with each of the three axes yields a filter with low-pass spatial and temporal frequency characteristics. Conversely, if $g(n)$ is convolved with each axis, the resulting filter will have band-pass spatial and temporal frequency characteristics. Additionally, since the separable 3D impulse responses for these two examples are identical in space and time, the transfer functions associated with the impulse responses will possess identical filter characteristics (e.g., bandwidth, transition region, cut-off frequency, center frequency) along each frequency axes. Thus, if the 3D frequency bandwidth of the discretely sampled input signal (i.e., the bandwidth of the signal projection onto the 0th approximation space) is contained in the volume shown in Figure 15a), convolving the rows, columns and frames with either $h(n)$ or $g(n)$ and downsampling by two yields discrete approximation and d^7 detail signals with the frequency supports shown in Figure 15b).

In order to obtain the remaining discrete detail signals $d_{-1:(l,m,n)}^1$ through $d_{-1:(l,m,n)}^6$, h and g are separately convolved in various combinations with each axis. The supporting regions in spatio-temporal frequency space of the resulting discrete detail signals are shown in Figure 16. The passband characteristics along each frequency axis for a given filter are determined by the order in which the spatio-temporal axes are convolved with h and g (see Equation 67). Like Mallat's 2D algorithm, the spatial frequency characteristics of the detail filters capture either horizontal, vertical or diagonal spatial frequency components in the scene (40). However, by adding a temporal dimension to the $L_2(\mathbb{R}^3)$ decomposition algorithm, one can capture these same spatial frequency components for either moving or stationary targets. Additionally, the multi-scale property of the decomposition algorithm generates spatio-temporal filters tuned to different object sizes and speeds. Thus, the $L_2(\mathbb{R}^3)$ discrete

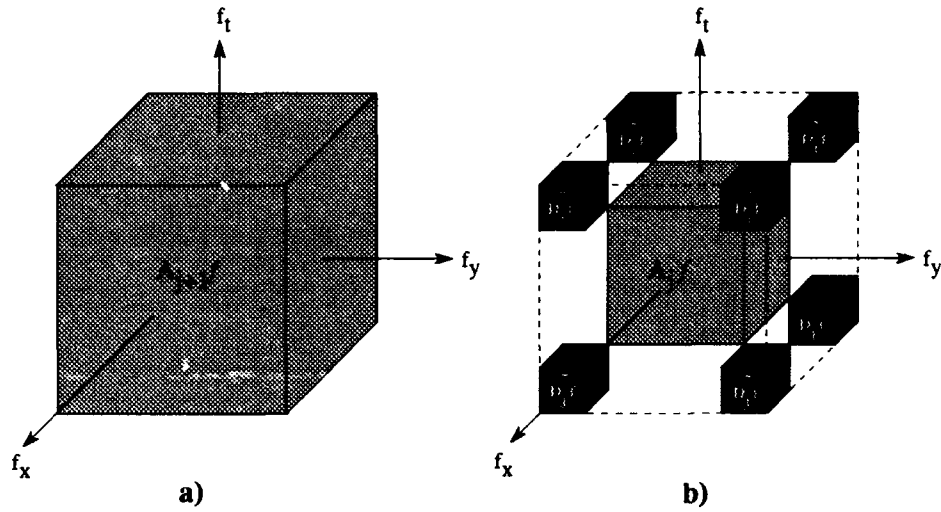


Figure 15. a) Spatio-temporal frequency volume of discretely sampled input signal, or, equivalently, the bandwidth of the $j = 0$ discrete approximation space. b) Frequency supports of the $j = -1$ approximation and detail spaces $A_{-1}f$ and $W_{-1}^T f$ respectively.

multiresolution analysis algorithm represents the decomposition of a 3D signal into a bank of independent *spatio-temporally* oriented frequency channels. The next section presents the results of applying the $L_2(\mathbb{R}^3)$ discrete multiresolution analysis algorithm to a synthetic scene consisting of a moving and a stationary object.

Finally, note that the $j = -1$ approximation region in Figure 15b) and the detail regions in Figure 16 combine to span the $j = 0$ approximation region in Figure 15a). If one were to implement the next stage of the sub-band decomposition algorithm, the $j = -1$ approximation frequency volume in Figure 15b) would be decomposed into constituent $j = -2$ approximation and detail frequency volumes identical to those contained in the $j = 0$ approximation volume but reduced in bandwidth *along each dimension* by a factor of two. The important point to note here is that each stage in the “conventional” $L_2(\mathbb{R}^3)$ decomposition process simultaneously reduces the spatial and temporal bandwidths of the filters. Thus, for any given spatial scale, one is forced to analyze the scene at the same temporal scale. This constraint precludes the possibility of analyzing multiple temporal scales (i.e. object speeds) for a fixed spatial scale, which in turn limits the tools effectiveness for the purpose of motion analysis (9). More will be said about this problem later in the chapter.

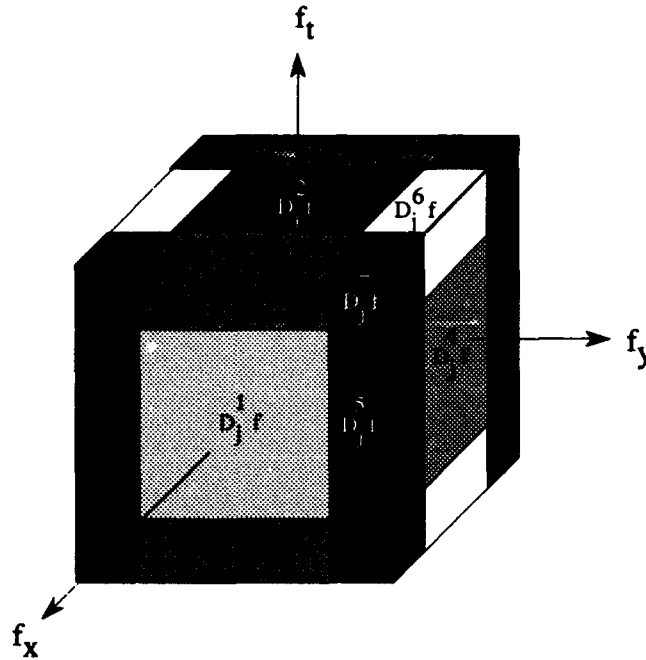


Figure 16. Spatio-temporal frequency volumes of discrete detail signal $d_{-1;(l,m,n)}^1$ through $d_{-1;(l,m,n)}^6$.

3.5 A Simple Application

The “conventional” $L_2(\mathbb{R}^3)$ discrete wavelet multiresolution analysis depicted by the oct-tree sub-band coding structure in Figure 14 is theoretically capable of extracting horizontal, vertical and diagonal features from moving or stationary objects in 3D imagery. In order to test these properties, the algorithm was applied to a $64 \times 64 \times 64$ synthetic image sequence. The image sequence was created on a Silicon Graphics computer and consists of a simple animated scene containing a stationary and a moving rectangle of equal size and intensity as shown in Figure 17. The moving rectangle starts in the upper left corner of the scene and moves to the lower right corner in a parabolic fashion, while the stationary rectangle remains fixed in the lower left corner. The sizes and speeds of the objects were constructed to prevent spatial or temporal aliasing (this topic is discussed further in Chapter IV).

The decomposition algorithm was written in C and implemented on several UNIX platforms including a NeXT, SUN SPARCstation 2, Silicon Graphics 4D, Silicon Graphics 8D, and a CRAY MPX. The discrete convolutions in Equation 67 were carried out with a three-dimensional shift-and-multiply routine rather than with filtering operations in the Fourier spatio-temporal frequency domain.

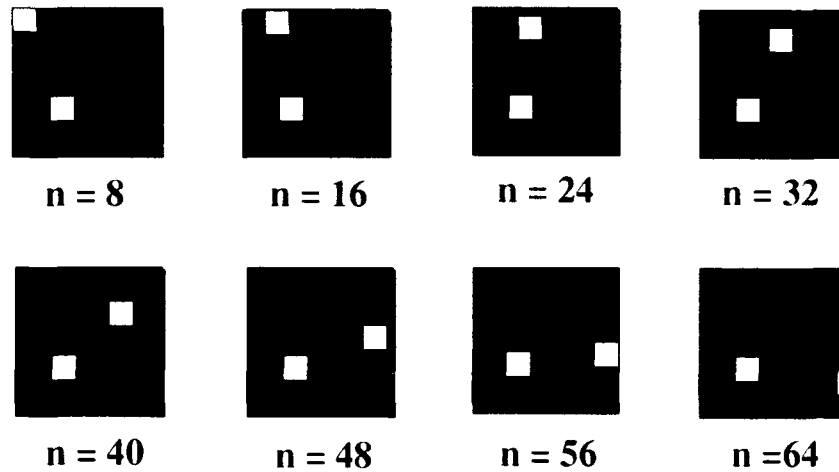


Figure 17. Several frames of an animated scene consisting of a stationary rectangle and a moving rectangle of equal size and intensity. Each frame contains 64×64 pixels.

Border problems, which are a common problem in convolution schemes, were reduced by making the borders symmetric about the spatial and temporal axes.

Assuming the discretely sampled image sequence represents the approximation coefficients at the resolution level $j = 0$, Figure 18 shows several frames containing the detail coefficients $d_{-1}^2, d_{-1}^3, d_{-1}^6$ and d_{-1}^7 , all of which were produced using a Daubechies 4 QMF pair in both space and time (14). Based on the frequency responses of the separable impulse responses in Equation 67, one expects that d_{-1}^2 and d_{-1}^3 will extract the horizontal features, and d_{-1}^6 and d_{-1}^7 will extract the diagonal features of either stationary or moving objects. It is instructive to compare these results with those obtained by applying a 2D multiresolution analysis to a simple rectangle as demonstrated by S. Mallat (Figure 19). Recall that under the 2D multiresolution analysis, an $L_2(\mathbb{R}^2)$ image is decomposed into a set of spatially oriented channels where each channel captures either vertical, horizontal or diagonal features of the image. On the other hand, Figure 18 shows that under an $L_2(\mathbb{R}^3)$ multiresolution analysis, the scene is decomposed into independent spatio-temporally oriented channels which now provide the ability to extract these same spatial details for either stationary or moving objects.

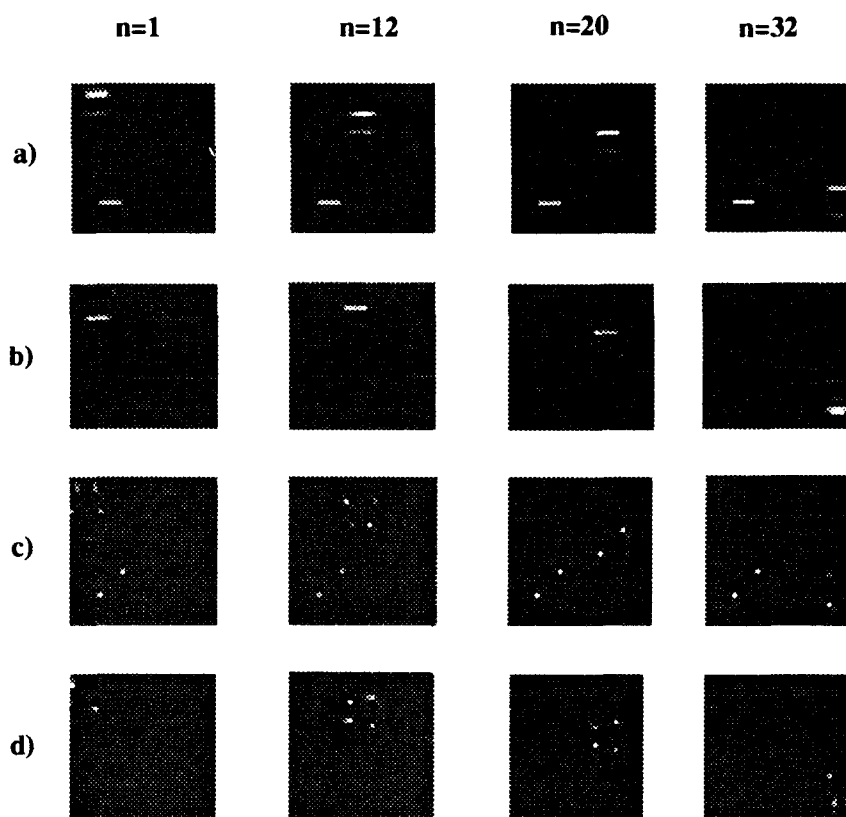


Figure 18. Four level $j = -1$ detail coefficients obtained by decomposing the scene in Figure 17 using a Daubechies 4 QMF pair in space and time. In a) and b) d_{-1}^2 and d_{-1}^3 respectively extract horizontal features of moving and/or stationary objects. In c) and d) d_{-1}^6 and d_{-1}^7 respectively extract diagonal features of stationary and/or moving objects.

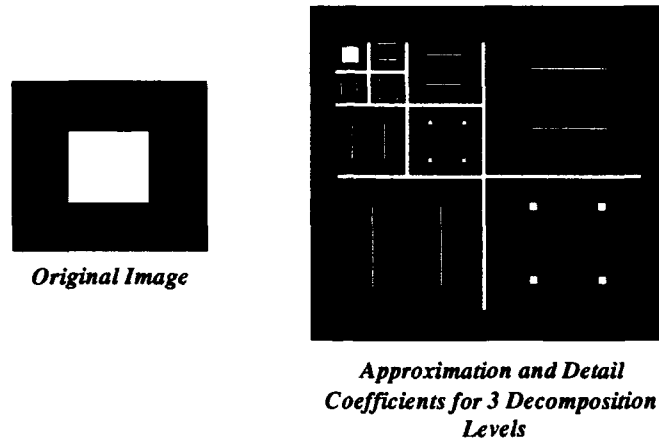


Figure 19. Multiscale results obtained by applying S. Mallat's 2D multiresolution analysis to a simple rectangle. These results were taken from (40). The figure depicts the ability of the 2D decomposition algorithm to extract horizontal, vertical and diagonal features at three different resolution. The square in the upper left hand corner is the final approximation signal produced by the decomposition process.

3.6 Discussion

The $L_2(\mathbb{R}^3)$ wavelet multiresolution analysis and the discrete decomposition algorithm presented here were constructed from an extension of Mallat's $L_2(\mathbb{R})$ and $L_2(\mathbb{R}^2)$ theory. Consequently, there are several limitations that carry over with the "conventional" extension that make it less than ideal for the analysis of motion. Three of these fundamental limitations are explained below.

In the conventional multiresolution analysis introduced by Meyer and Mallat, the 2D approximation space $L_2(\mathbb{R}^2)$ was created from two identical 1D approximation spaces. This generates an approximation, or scaling function, filter with identical frequency characteristics in the f_x and f_y spatial frequency dimensions. Similarly, the $L_2(\mathbb{R}^3)$ approximation space developed in this chapter was formed from the tensor product of three identical $L_2(\mathbb{R})$ approximation spaces. Like the separable 2D wavelet filter, the corresponding 3D filter has identical passband, stopband and transition region characteristics in f_x , f_y and f_t . The major drawback to this approach is that it does not provide the flexibility to tailor the spatio-temporal frequency characteristics of the wavelet filter to match the frequency behavior of the 3D signal under analysis. For example, a particular problem may require a narrow transition region between the temporal frequency passband and stopband, but allow a much wider spatial frequency transition region. Since a wider transition region can be obtained with a lower

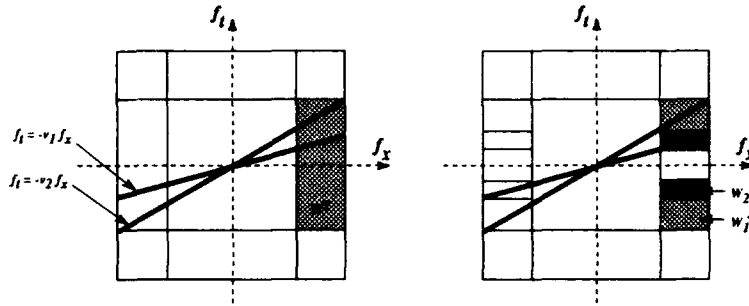


Figure 20. a) Black lines represent the frequency support of two identical objects moving at different velocities superimposed on the filters formed in frequency space by one step in a conventional multiresolution decomposition. b) Frequency support of moving objects superimposed on filters formed by one spatial and two temporal decompositions. Notice how the filters now separate the two objects.

order filter (i.e., fewer h and g coefficients), a multiresolution analysis constructed from filters with non-homogeneous spatial and temporal frequency characteristics would improve the computational efficiency of the design. In the conventional $L_2(\mathbb{R}^3)$ wavelet multiresolution analysis, the designer must use an identical higher-order filter in both the spatial and temporal frequency dimensions in order to meet the temporal frequency design specifications.

A second important limitation of the conventional $L_2(\mathbb{R}^3)$ multiresolution analysis is that it restricts the analysis of spatial and temporal details in an image sequence to the same resolution level (9). In order to demonstrate how this poses a problem for a spatio-temporal frequency motion analysis approach, consider the 2D motion problem of two identical 1D rectangles moving with slightly different velocities. Assume that the size and intensity of the rectangles remain constant in time, and that they move with the constant translational velocity components v_1 and v_2 . As discussed in Chapter II, the Fourier transform of the moving rectangles are given by $F(f_x) \cdot \delta(f_t + v_1 f_x)$ and $F(f_x) \cdot \delta(f_t + v_2 f_x)$. Thus, constant velocity motion in one spatial dimension shifts the Fourier transform of the stationary rectangles, $F(f_x)$, to two lines in 2D frequency space defined by $f_t = -v_1 f_x$ and $f_t = -v_2 f_x$. This behavior is demonstrated in Figure 20a) where the lines represent the regions of support of the moving rectangle's Fourier transforms. It is assumed here that the magnitude of v_1 is slightly less than v_2 . The lines are superimposed on the wavelet filterbank formed by one step in a conventional multiresolution wavelet decomposition.

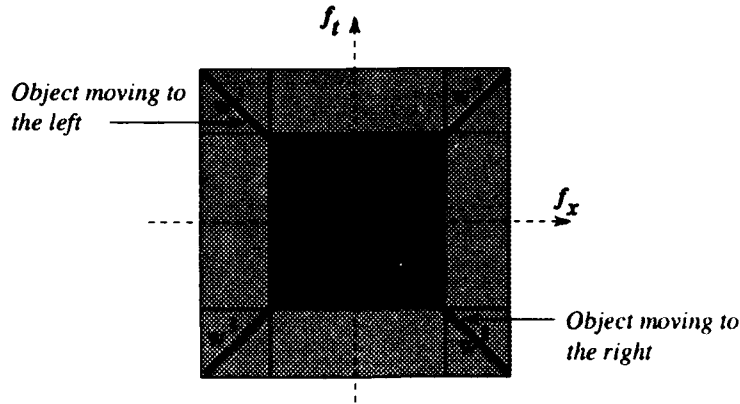


Figure 21. Black lines represent the frequency supports of two 1D objects moving at the same speed in opposite directions superimposed on the annulus formed in the Fourier plane by two steps in a conventional multiresolution decomposition.

In order to discriminate between the two moving objects, spatio-temporal frequency based motion analysis techniques commonly employ filters designed to separate their spectrums in frequency space. However, in example a), the shaded filter marked W^2 generated by the conventional multiresolution decomposition cannot resolve the two spectrums. Now consider the frequency support superimposed on the “unconventional” filter bank shown in Figure 20b). These filters, marked W_1^2 and W_2^2 , were produced by decomposing the signal once in space and *twice* in time. Thus, by “decoupling” the spatial and temporal decomposition stages of the conventional discrete wavelet multiresolution analysis, one can clearly resolve the two spectra. This “decoupling” process is described in more detail in Chapter IV.

The third problem with using a conventional multiresolution wavelet theory for motion analysis is that it is not directionally selective. For example, now consider a pair of 1D moving rectangles, one of which moves to the right at a speed v and the other which moves to the left at the same speed. Their frequency supports are shown by the crossed lines in Figure 21. Notice that both lines pass through the frequency support of the detail space W^3 . Since the 2D wavelet associated with this space is real, its Fourier transform is symmetric about both the f_x and f_y axes. Thus, the W^3 wavelet will respond equally to an object moving in either direction at the speed v . The conventional discrete wavelet transform therefore is clearly not directionally selective. Chapter IV presents a solution to this problem by incorporating a Hilbert Transform into the $L_2(\mathbb{R}^3)$ wavelet multiresolution analysis.

3.7 Conclusion

Y. Meyer's theory for wavelet multiresolution analyses in $L_2(\mathbb{R}^n)$ does not provide details for constructing orthonormal wavelet bases for $L_2(\mathbb{R}^3)$. Furthermore, previous instantiations of Meyer's wavelet multiresolution analysis dealt exclusively with $L_2(\mathbb{R})$ and $L_2(\mathbb{R}^2)$ signals (14, 39). Thus, the first section of this chapter provided the mathematical details for the construction of wavelet orthonormal bases for the space of finite energy spatio-temporal signals, $L_2(\mathbb{R}^3)$. Theorem 1 shows this basis set consists of seven dyadically dilated and translated wavelets which represent "independent" spatio-temporal channels in 3D Fourier frequency space. In the second section, an "oct-tree" sub-band coding scheme was presented for implementing the Discrete Spatio-temporal Wavelet Transform. The algorithm generates a bank of octave-band filters such that each filter possesses uniform spatial and temporal frequency characteristics. The sub-band decomposition algorithm was applied to a set of synthetic 3D imagery to demonstrate its ability to extract vertical, horizontal or diagonal features from moving or stationary targets. Lastly, three important problems were described which limit the utility of the conventional wavelet multiresolution decomposition algorithm for motion analysis. These problems are resolved in the following chapter.

IV. A Non-Homogeneous, Motion-Oriented $L_2(\mathbb{R}^3)$ Wavelet Multiresolution Analysis

4.1 Introduction.

The previous chapter discussed three major problems associated with using a conventional $L_2(\mathbb{R}^3)$ wavelet multiresolution analysis for segmenting and characterizing moving objects in time-sequential imagery. The purpose of this chapter is to present solutions for two of these problems. The two problems concern 1) the restrictions placed on the 3D wavelet filter design process by the theoretical development of the *homogeneous* approximation space V_j , and 2) the oct-tree decomposition architecture that constrains the analysis of spatial and temporal details to the same resolution in space and time. The first problem is addressed in the following section, where it is shown that an $L_2(\mathbb{R}^3)$ wavelet multiresolution analysis can be constructed from a separable scaling function formed from three non-identical $L_2(\mathbb{R})$ scaling functions. This provides the flexibility to build wavelet filters with non-homogeneous spatial and temporal frequency characteristics. Next a solution to the second problem is presented which essentially “decouples” the spatial and temporal decomposition processes using a modified wavelet packet and a non-standard decomposition tree structure. The resulting algorithm, referred to as a motion-oriented wavelet multiresolution analysis, yields an analytical tool with independent zoom-in and zoom-out capabilities in space and time. Since the algorithm is discrete in both space and time, one must consider how it is affected by spatial and temporal aliasing. This problem is examined in the fourth section. In particular, the aliasing problem is addressed in terms of its affect on the discretely sampled input signal. The last major section of the chapter presents several results obtained by applying the non-homogeneous, motion-oriented wavelet multiresolution analysis to different sequences of synthetic and real IR imagery. The chapter concludes by summarizing the capabilities and limitations of the new motion analysis tool.

4.2 A Non-Homogeneous Wavelet Multiresolution Analysis for $L_2(\mathbb{R}^3)$.

The approximation space V_j of the conventional $L_2(\mathbb{R}^3)$ wavelet multiresolution analysis presented in the previous chapter was formed from the tensor product of three identical approximation spaces. This approach produced a scaling function filter with identical passband characteristics in f_x , f_y and f_t . This in turn limits the filter designer’s ability to tailor the spatial and temporal frequency characteristics of the wavelet filter to match the frequency behavior of the 3D signal under analysis.

This section demonstrates one can increase the flexibility of the design process by creating a “non-homogeneous” wavelet multiresolution analysis for $L_2(\mathbb{R}^3)$ signals from non-identical spatial and temporal $L_2(\mathbb{R})$ approximation spaces. The section begins by creating an $L_2(\mathbb{R}^3)$ approximation spaces from a separable 3D scaling function formed by multiplying different 1D spatial and temporal scaling functions.

4.2.1 Separable Scaling Function and Approximation Space. Let ϕ be a scaling function such that $\{2^{\frac{j}{2}}\phi(2^j \cdot -n) \mid n \in \mathbb{Z}\}$ forms an orthonormal basis for the multiresolution approximation, V_j , of $L_2(\mathbb{R})$. Let $\tilde{\phi}$ be a different scaling function such that $\{2^{\frac{j}{2}}\tilde{\phi}(2^j \cdot -n) \mid n \in \mathbb{Z}\}$ forms an orthonormal basis for the multiresolution approximation, $\tilde{V}_j$, of $L_2(\mathbb{R})$. Define the separable, closed, linear subspaces of $L_2(\mathbb{R}^3)$ by

$$V_j = V_j^x \otimes V_j^y \otimes \tilde{V}_j^t = \overline{\text{Span}\{F(x, y, t) = f(x)g(y)h(t) \mid f \in V_j^x, g \in V_j^y \text{ and } h \in \tilde{V}_j^t\}} \quad (68)$$

Given the above definition of the approximation space V_j , Theorem 2 shows there exists a separable 3D scaling function such that the set comprised of all its integer translations forms an orthonormal basis for V_j .

Theorem 2. For each $j \in \mathbb{Z}$, the set of functions $\{2^{\frac{3j}{2}}\phi(2^j x - l)\phi(2^j y - m)\tilde{\phi}(2^j t - n) \mid (l, m, n) \in \mathbb{Z}^3\}$ forms an orthonormal basis for V_j .

Proof Let $\Phi_{j:(l,m,n)}(x, y, t) = 2^{\frac{3j}{2}}\phi(2^j x - l)\phi(2^j y - m)\tilde{\phi}(2^j t - n)$. Then

$$\langle \Phi_{j:(l,m,n)}, \Phi_{j:(l',m',n')} \rangle = \begin{cases} 1 & \text{if } l = l' \text{ and } m = m' \text{ and } n = n' \\ 0 & \text{otherwise} \end{cases} \quad (69)$$

where $\langle \cdot, \cdot \rangle$ denotes the inner product on $L_2(\mathbb{R}^3)$. Equation 69 implies the set of vectors $\{\Phi_{j:(l,m,n)} \mid (l, m, n) \in \mathbb{Z}^3\}$ forms an orthonormal set in $L_2(\mathbb{R}^3)$.

Now, let F be a vector in V_j . By definition of V_j , $F(x, y, t) = f(x)g(y)h(t)$ for some $f \in V_j^x$, $g \in V_j^y$ and $h \in \tilde{V}_j^t$. Expressing f , g and h in terms of their respective orthonormal bases and rearranging terms yields

$$F = \sum_l 2^j \langle f, \phi_{j:l} \rangle \phi_{j:l} \sum_m 2^j \langle g, \phi_{j:m} \rangle \phi_{j:m} \sum_n 2^j \langle h, \tilde{\phi}_{j:n} \rangle \tilde{\phi}_{j:n}$$

$$\begin{aligned}
&= \sum_l \sum_m \sum_n 2^{\frac{3j}{2}} \langle fgh, \phi_{j;l} \phi_{j;m} \tilde{\phi}_{j;n} \rangle \phi_{j;l} \phi_{j;m} \tilde{\phi}_{j;n} \\
&= \sum_l \sum_m \sum_n 2^{\frac{3j}{2}} \langle F, \Phi_{j;(l,m,n)} \rangle \Phi_{j;(l,m,n)}
\end{aligned} \tag{70}$$

where $\phi_{j;q} = \phi(2^j \cdot -q)$ and $\tilde{\phi}_{j;q} = \tilde{\phi}(2^j \cdot -q)$. Equation 70 shows F can be expressed as a Fourier series expansion of the orthonormal set $\{\Phi_{j;(l,m,n)} \mid (l,m,n) \in \mathbb{Z}^3\}$. Thus, by the Fourier Series Theorem (45), $\{\Phi_{j;(l,m,n)} \mid (l,m,n) \in \mathbb{Z}^3\}$ forms an orthonormal basis for V_j . **Q.E.D.**

Theorem 2 allows one to create an $L_2(\mathbb{R}^3)$ approximation space from non-identical $L_2(\mathbb{R})$ approximation spaces. Furthermore, it shows that the resulting approximation space is spanned by integer translations of a separable scaling function formed from the product of three non-identical scaling functions. The next section proves that the scaling function and approximation space generate a multiresolution analysis. The multiresolution analysis is then used in section 4.2.3 to construct an orthonormal wavelet basis for $L_2(\mathbb{R}^3)$ comprised of wavelets with non-homogeneous spatio-temporal frequency characteristics.

4.2.2 Multiresolution Analysis. In order to construct a multiresolution analysis, recall from Section 2.2.3 that the approximation spaces V_j must possess the following properties: there must 1) exist a chain of closed linear spaces V_j ,

$$\cdots V_{-2} \subset V_{-1} \subset V_{-0} \subset V_1 \subset V_2 \subset \cdots \tag{71}$$

such that 2)

$$\overline{\bigcup_{j \in \mathbb{Z}} V_j} = L_2(\mathbb{R}^3) \quad \text{and} \quad \bigcap_{j \in \mathbb{Z}} V_j = \{0\} \tag{72}$$

and where 3)

$$\begin{aligned}
f(x, y, t) \in V_j &\Leftrightarrow f(2x, 2y, 2t) \in V_{j+1}; \quad j \in \mathbb{Z} \\
f(x, y, t) \in V_j &\Rightarrow f\left(x + \frac{l}{2^j}, y + \frac{m}{2^j}, t + \frac{n}{2^j}\right) \in V_j; \quad (l, m, n) \in \mathbb{Z}^3
\end{aligned} \tag{73}$$

Theorem 3 shows the non-homogeneous approximation spaces V_j do indeed satisfy these properties.

Theorem 3. The family of closed, linear spaces, $\{V_j \mid j \in \mathbb{Z}\}$, forms a multiresolution analysis in $L_2(\mathbb{R}^3)$.

Proof. To prove property 1), it will suffice to show $V_j \subset V_{j+1}$ for arbitrary $j \in \mathbb{Z}$. Let $F \in V_j$ where $F(x, y, t) = f(x)g(y)h(t)$ such that $f \in V_j^x, g \in V_j^y$ and $h \in \tilde{V}_j^t$. Now, $f \in V_j^x, g \in V_j^y$ and $h \in \tilde{V}_j^t$ implies $f \in V_{j+1}^x, g \in V_{j+1}^y$ and $h \in \tilde{V}_{j+1}^t$. But,

$$V_{j+1} = \overline{\text{Span}\{u(x)v(y)w(t) \mid u \in V_{j+1}^x, v \in V_{j+1}^y \text{ and } w \in \tilde{V}_{j+1}^t\}}$$

Thus, the vector F must also be contained in V_{j+1} , implying $V_j \subset V_{j+1}$.

To prove the denseness condition of property 2), let

$$M = \overline{\bigcup_{j \in \mathbb{Z}} V_j} \quad (74)$$

and assume M is not equal to $L_2(\mathbb{R}^3)$. M is therefore a proper subspace of $L_2(\mathbb{R}^3)$ and, by Hahn-Banach (45), there exists a linear functional ℓ on $L_2(\mathbb{R}^3)$ such that $\ell(M) = 0 \forall M \in M$ and $\ell(G) \neq 0$ for some $G \in L_2(\mathbb{R}^3) - M$. Then, by the Riesz Representation Theorem (54), there exists a unique $H \in L_2(\mathbb{R}^3)$ such that

$$\ell(F) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F(x, y, t) H(x, y, t) dx dy dt \quad (75)$$

$\forall F \in L_2(\mathbb{R}^3)$. Furthermore, if ℓ does not equal the zero functional, then $H \neq 0$. Additionally, $\ell(M) = 0 \forall M \in M$ implies $H \perp M$. Consequently, the orthogonal projection of H onto $V_j \in M$, $P_j H$, must equal zero. Now, since $H \in L_2(\mathbb{R}^3)$, there exists a compactly supported C^∞ function, H_o , such that $\|H_o - H\| < \epsilon$. And, by the Orthogonal Projection Theorem, $\|P_j H_o\| = \|P_j(H_o - H)\| \leq \|H_o - H\| < \epsilon$. Thus, by Parseval's Identity,

$$\begin{aligned} \|P_j H_o\|^2 &= 2^{\frac{3j}{2}} \sum_l \sum_m \sum_n |\langle \phi(2^j x - l) \phi(2^j y - m) \tilde{\phi}(2^j t - n), H_o(x, y, t) \rangle|^2 \\ &\leq \epsilon^2 \end{aligned} \quad (76)$$

Using standard mathematical manipulations, it can be shown that

$$\sum_l \sum_m \sum_n |\langle \phi(2^j x - l) \phi(2^j y - m) \tilde{\phi}(2^j t - n), H_o(x, y, t) \rangle|^2 = \quad (77)$$

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |\hat{H}_o(\xi, \eta, \tau)|^2 |\hat{\phi}(2^{-j}\xi) \hat{\phi}(2^{-j}\eta) \hat{\phi}(2^{-j}\tau)|^2 d\xi d\eta d\tau + R_j \quad (78)$$

where $\hat{H}_o$ denotes the Fourier Transform of H_o and

$$|R_j| \leq \sum_{l \neq 0} \sum_{m \neq 0} \sum_{n \neq 0} |\hat{H}_o(\xi + 2\pi 2^j l, \eta + 2\pi 2^j m, \tau + 2\pi 2^j n)| |\hat{f}(\xi, \eta, \tau)| \quad (79)$$

Now consider the sequence of functions

$$H_j(\xi, \eta, \tau) = \sum_{l \neq 0} \sum_{m \neq 0} \sum_{n \neq 0} |\hat{H}_o(\xi + 2\pi 2^j l, \eta + 2\pi 2^j m, \tau + 2\pi 2^j n)| \quad (80)$$

Since H_o is a compact C^∞ function, $\hat{H}_o$ is uniformly bounded and $H_j \rightarrow 0$ as $j \rightarrow \infty$. Additionally, $\hat{H}_o \in L_1(\mathbb{R}^3)$ implies $R_j \rightarrow 0$ as $j \rightarrow \infty$ (11). Moreover, $\hat{\phi}$ and $\hat{\tilde{\phi}}$ are continuous and uniformly bounded and $\hat{\phi}(0) = \hat{\tilde{\phi}}(0) = 1$. Hence, Lebesgue's Dominated Convergence Theorem can be applied in conjunction with Equation 76 to obtain

$$\lim_{j \rightarrow \infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |\hat{H}_o(\xi, \eta, \tau)|^2 |\hat{\phi}(2^{-j}\xi) \hat{\phi}(2^{-j}\eta) \hat{\phi}(2^{-j}\tau)|^2 d\xi d\eta d\tau = \|H_o\|^2 \leq \epsilon^2 \quad (81)$$

Finally, $\|H_o\| \leq \epsilon$ and $\|H_o - H\| < \epsilon$ implies $\|H\| \leq 2\epsilon$. But ϵ arbitrarily small implies $H = 0$, which contradicts our original assumption. Thus, M is dense in $L_2(\mathbb{R}^3)$.

To prove the intersection property of Proposition 2) let

$$M = \bigcap_{j \in \mathbb{Z}} V_j \quad (82)$$

Since each element V_j in M is closed, M is closed and $\{0\}$ is therefore clearly contained in M . Now, let $F(x, y, t) = f(x)g(y)h(t)$ be an element in M . Then, F is contained in V_j for all $j \in \mathbb{Z}$. And, by the definition of V_j , $f \in V_j^x$, $g \in V_j^y$ and $h \in V_j^t$ for all $j \in \mathbb{Z}$. But, the sequence of spaces $\{V_j \mid j \in \mathbb{Z}\}$ forms a multiresolution analysis for $L_2(\mathbb{R})$, implying $f = 0$, $g = 0$ and $h = 0$. Thus, F is contained in $\{0\}$ and $M = \{0\}$.

Like Property 1), Property 3) follows easily from the fact that V_j is constructed from a tensor product of three multiresolution approximations of $L_2(\mathbb{R})$. If $F(x, y, t) = f(x)g(y)h(t) \in V_j$ then $f(x) \in V_j^x$, $g(y) \in V_j^y$ and $h(t) \in \tilde{V}_j^t$. Thus, $f(2x) \in V_{j+1}^x$, $g(2y) \in V_{j+1}^y$ and $h(2t) \in \tilde{V}_{j+1}^t$.

implying $F(2x, 2y, 2t) \in V_{j+1}$. Moreover, $f(x) \in V_j^x$, $g(y) \in V_j^y$ and $h(t) \in \tilde{V}_j^t$ implies $f(x + \frac{l}{2^j}) \in V_j^x$, $g(y + \frac{m}{2^j}) \in V_j^y$ and $h(t + \frac{n}{2^j}) \in \tilde{V}_j^t$. Thus, $F(x + \frac{l}{2^j}, y + \frac{m}{2^j}, t + \frac{n}{2^j}) \in V_j$. **Q.E.D.**

Theorem 3 proves one can construct a multiresolution analysis from the non-homogeneous separable scaling function $\phi(x)\phi(y)\tilde{\phi}(t)$. Since the existence of an orthonormal wavelet basis is guaranteed by the formation of a multiresolution analysis (39, 42), the purpose of the next section is to describe the properties of such a basis set. In particular, it will be shown that the 3D wavelets in the basis set are formed from the product of three non-identical 1D functions, allowing one to independently control the spatial and temporal frequency characteristics of the wavelet filters during the filter design process.

4.2.3 Orthonormal Wavelet Basis. In the 3D multiresolution analyses, approximations of a spatio-temporal signal at the j th and $(j+1)$ st resolutions in space and time are obtained by orthogonally projecting the signal respectively onto the spaces V_j and V_{j+1} . The spatial and temporal details that comprise the difference in information between these two approximations are then contained in the orthogonal complement of V_j in V_{j+1} . As in Chapter 3, this complementary space is denoted by the symbol W_j . Theorem 4 shows an orthonormal basis for W_j (and for $L_2(\mathbb{R}^3)$) consists of seven sets of scaled and translated "wavelets."

Theorem 4. Let ψ and $\tilde{\psi}$ be the one-dimensional wavelets respectively generated by the scaling functions ϕ and $\tilde{\phi}$. Then the seven "wavelets"

$$\begin{aligned}\Psi_j^1(x, y, t) &= 2^{\frac{3j}{2}} \phi(2^j x) \phi(2^j y) \tilde{\psi}(2^j t) \\ \Psi_j^2(x, y, t) &= 2^{\frac{3j}{2}} \phi(2^j x) \psi(2^j y) \tilde{\phi}(2^j t) \\ \Psi_j^3(x, y, t) &= 2^{\frac{3j}{2}} \phi(2^j x) \psi(2^j y) \tilde{\psi}(2^j t) \\ \Psi_j^4(x, y, t) &= 2^{\frac{3j}{2}} \psi(2^j x) \phi(2^j y) \tilde{\phi}(2^j t) \\ \Psi_j^5(x, y, t) &= 2^{\frac{3j}{2}} \psi(2^j x) \phi(2^j y) \tilde{\psi}(2^j t) \\ \Psi_j^6(x, y, t) &= 2^{\frac{3j}{2}} \psi(2^j x) \psi(2^j y) \tilde{\phi}(2^j t) \\ \Psi_j^7(x, y, t) &= 2^{\frac{3j}{2}} \psi(2^j x) \psi(2^j y) \tilde{\psi}(2^j t)\end{aligned}\tag{83}$$

are such that for each $j \in \mathbb{Z}$, $\{\Psi_j^p(x-l, y-m, t-n) \mid (l, m, n) \in \mathbb{Z}^3; p = 1, 2, \dots, 7\}$ forms an orthonormal basis for W_j ; and $\{\Psi_j^p(x-l, y-m, t-n) \mid j \in \mathbb{Z}; (l, m, n) \in \mathbb{Z}^3; p = 1, 2, \dots, 7\}$ forms an orthonormal basis for $L_2(\mathbb{R}^3)$.

The proof of Theorem 4 is not provided here because, with one exception, it follows precisely the proof of Theorem 1 contained in Section 3.2. The only difference between the two proofs is that the time dependent scaling function and wavelet $\phi(t)$ and $\psi(t)$ in the proof of Theorem 1 are now replaced everywhere by the new (and different) functions $\tilde{\phi}(t)$ and $\tilde{\psi}(t)$ respectively. Because the orthonormal wavelet basis for a given detail space is formed from the product of three non-identical 1D spatial and temporal wavelets, the resulting wavelet filter for that space has non-homogeneous spatial and temporal frequency characteristics. Also, because the wavelet basis for each detail space is separable in space and time, the filter designer can easily and independently control the spatial and temporal frequency behavior of the wavelet filter. This property will prove valuable in the following chapter where the spatial and temporal frequency characteristics of the 3D filter are adapted to match the spatial and velocity behavior of a moving object.

In the next section, discrete versions of the non-homogeneous wavelet filters are used in an extension of Mallat's 2D discrete multiresolution analysis referred to here as a "non-homogeneous $L_2(\mathbb{R}^3)$ discrete wavelet multiresolution analysis." Since the development parallels the construction of the homogeneous $L_2(\mathbb{R}^3)$ discrete multiresolution analysis described in Chapter III, many of the details are left to the reader. Also, as was the case with the homogeneous discrete wavelet multiresolution analysis, the resulting non-homogeneous oct-tree decomposition structure is somewhat impractical for the analysis of moving objects. Thus, the non-homogeneous oct-tree structure is presented here more for completeness than for its intended use as a motion analysis tool. A non-conventional discrete decomposition structure more suited to the analysis of motion is presented later in the chapter.

4.2.4 Discrete Multiresolution Decomposition Algorithm. The development of the discrete, non-homogeneous wavelet multiresolution decomposition algorithm closely follows the derivation in Section 3.3 for the homogeneous case. This section, therefore, will briefly present a derivation of one branch of the oct-tree decomposition algorithm, and simply list the discrete convolution operations that comprise the remaining seven branches.

Like the derivation in Chapter III, begin by assuming the sequence obtained by sampling the signal in x, y and t represents the coefficients associated with the orthogonal projection of f onto the approximation space V_o . Since the chain of approximation spaces $\{V_j \mid j \in \mathbb{Z}\}$ forms a multiresolution analysis in $L_2(\mathbb{R}^3)$, any basis element in V_j can be expanded in terms of the basis elements of V_{j+1} . Therefore, given the basis element $2^{\frac{j}{2}}\phi(2^j x - l)\phi(2^j y - m)\tilde{\phi}(2^j t - n)$ in V_j one can write

$$\begin{aligned} \phi(2^j x - l)\phi(2^j y - m)\tilde{\phi}(2^j t - n) &= 2^{3(j+1)} \sum_p \sum_q \sum_r \langle \phi(2^j u - l)\phi(2^j v - m)\tilde{\phi}(2^j w - n), \\ &\phi(2^{j+1} u - p)\phi(2^{j+1} v - q)\tilde{\phi}(2^{j+1} w - r) \rangle \phi(2^{j+1} x - p)\phi(2^{j+1} y - q)\tilde{\phi}(2^{j+1} t - r) \end{aligned} \quad (84)$$

where $p, q, r \in \mathbb{Z}$. Expanding the inner product in the above expression in its integral form yields

$$\begin{aligned} 2^{3(j+1)} \langle \cdot, \cdot \rangle &= 2^{3(j+1)} \int \int \int_{-\infty}^{\infty} [\phi(2^j u - l)\phi(2^j v - m)\tilde{\phi}(2^j w - n)] \\ &\cdot [\phi(2^{j+1} u - p)\phi(2^{j+1} v - q)\tilde{\phi}(2^{j+1} w - r)] du dv dw \end{aligned} \quad (85)$$

Again, using the variable substitutions

$$\begin{aligned} \frac{a}{2} &= 2^j u - l \\ \frac{b}{2} &= 2^j v - m \\ \frac{c}{2} &= 2^j w - n \end{aligned} \quad (86)$$

the right hand side of Equation 85 can be rewritten as

$$\int \int \int_{-\infty}^{\infty} [\phi(\frac{a}{2})\phi(\frac{b}{2})\tilde{\phi}(\frac{c}{2})][\phi(a - (p - 2l))\phi(b - (q - 2m))\tilde{\phi}(c - (r - 2n))]\phi(a - (p - 2l))\phi(b - (q - 2m))\tilde{\phi}(c - (r - 2n)) da db dc \quad (87)$$

Next, defining the functions h and $\tilde{h}$ by

$$\begin{aligned} h(x) &\equiv \int_{-\infty}^{\infty} \phi(\frac{y}{2})\phi(y - x) dy \\ \tilde{h}(x) &\equiv \int_{-\infty}^{\infty} \tilde{\phi}(\frac{y}{2})\tilde{\phi}(y - x) dy \end{aligned} \quad (88)$$

the right hand side of Equation 85 can be combined with the "impulse responses" in Equation 88 to yield the following expression for the arbitrary basis element $2^{\frac{3j}{2}} \phi(2^j x - l) \phi(2^j y - m) \tilde{\phi}(2^j t - n)$

$$2^{\frac{3j}{2}} \phi(2^j x - l) \phi(2^j y - m) \tilde{\phi}(2^j t - n) = \sum_p \sum_q \sum_r \bar{h}(2l - p) \bar{h}(2m - q) \bar{\tilde{h}}(2n - r) \phi(2^{j+1} x - p) \phi(2^{j+1} y - q) \tilde{\phi}(2^{j+1} t - r) \quad (89)$$

where $\bar{h}(x) = h(-x)$ and $\bar{\tilde{h}}(x) = \tilde{h}(-x)$. In order to obtain the coefficients associated with the signal projection onto V_j , one next forms the inner product of f with the arbitrary basis element in Equation 89 as follows

$$\begin{aligned} a_{j;l,m,n} &= \langle f, 2^{\frac{3j}{2}} \phi(2^j x - l) \phi(2^j y - m) \tilde{\phi}(2^j t - n) \rangle \\ &= \sum_p \sum_q \sum_r \bar{h}(2l - p) \bar{h}(2m - q) \bar{\tilde{h}}(2n - r) \langle f, \phi(2^{j+1} x - p) \phi(2^{j+1} y - q) \tilde{\phi}(2^{j+1} t - r) \rangle \\ &= \sum_p \sum_q \sum_r \bar{h}(2l - p) \bar{h}(2m - q) \bar{\tilde{h}}(2n - r) a_{j+1;p,q,r} \\ &= [a_{j+1;p,q,r} * \bar{h}(p) * \bar{h}(q) * \bar{\tilde{h}}(r)] (2l, 2m, 2n) \end{aligned} \quad (90)$$

where "*" is the discrete convolution operator. Equation 90 shows that the discrete representation of the orthogonal projection of the signal onto the approximation space V_j is obtained by discretely convolving the coefficients of the projection onto the next higher resolution level, V_{j+1} , with the separable, non-homogeneous impulse response $h(-p)h(-q)\tilde{h}(-r)$ and keeping every other sample in each dimension. Following a similar procedure, the seven discrete 3D convolution operations that produce the coefficients of the orthogonal projection onto the detail space W_j are given by

$$\begin{aligned} d_{j;l,m,n}^1 &= [a_{j+1;p,q,r} * \bar{h}(p) * \bar{h}(q) * \bar{\tilde{g}}(r)] (2l, 2m, 2n) \\ d_{j;l,m,n}^2 &= [a_{j+1;p,q,r} * \bar{h}(p) * \bar{g}(q) * \bar{\tilde{h}}(r)] (2l, 2m, 2n) \\ d_{j;l,m,n}^3 &= [a_{j+1;p,q,r} * \bar{h}(p) * \bar{g}(q) * \bar{\tilde{g}}(r)] (2l, 2m, 2n) \\ d_{j;l,m,n}^4 &= [a_{j+1;p,q,r} * \bar{g}(p) * \bar{h}(q) * \bar{\tilde{h}}(r)] (2l, 2m, 2n) \\ d_{j;l,m,n}^5 &= [a_{j+1;p,q,r} * \bar{g}(p) * \bar{h}(q) * \bar{\tilde{g}}(r)] (2l, 2m, 2n) \\ d_{j;l,m,n}^6 &= [a_{j+1;p,q,r} * \bar{g}(p) * \bar{g}(q) * \bar{\tilde{h}}(r)] (2l, 2m, 2n) \\ d_{j;l,m,n}^7 &= [a_{j+1;p,q,r} * \bar{g}(p) * \bar{g}(q) * \bar{\tilde{g}}(r)] (2l, 2m, 2n) \end{aligned} \quad (91)$$

where $\bar{g}(n) = \tilde{g}(-n)$ and $h(n), g(n)$ and $\tilde{h}(n), \tilde{g}(n)$ represent two different QMF pairs. Equations 90 and 91 form all eight branches of the non-homogeneous oct-tree decomposition structure shown in Figure 22. As before, the detail signals are obtained by convolving the discrete approximation signal at the next higher resolution level along each axis with various combinations the discrete "impulse responses" $h(n), g(n)$ and $\tilde{h}(n), \tilde{g}(n)$. The convolutions can also be viewed as filtering operations in discrete 3D frequency space, where the separable 3D filters are constructed from non-identical 1D spatial and temporal filters. The discrete filter bank designer can now quickly and easily combine different spatial and temporal QMF pairs to match the spatial and temporal frequency characteristics of the signal. For example, one can construct a discrete 3D filter using, say, a Daubechies 4 QMF pair for the spatial convolutions and a Daubechies 9 QMF pair for the temporal convolution. This yields a filter with a larger passband and a narrower transition region along the temporal frequency axis than along the spatial frequency axes. The design trade-off, of course, is that in order to meet the "tighter" temporal design requirements, one must use a higher order filter. The practicality of this design flexibility will be more evident in the following section where an unconventional decomposition algorithm is presented which allows one to examine an image sequence at multiple resolutions in time for a fixed resolution in space.

4.3 A Motion-Oriented Wavelet Multiresolution Analysis for $L_2(\mathbb{R}^3)$

In the discussion section at the end of Chapter 3, an argument was made against using the conventional oct-tree decomposition structure for analyzing motion in an image sequence. Essentially, it was shown that the oct-tree structure generates a filter bank comprised of analysis filters whose spatial and temporal bandwidths both decrease equally by a factor of two from one stage of the decomposition to the next. That is, it does not allow one to simultaneously examine the image sequence at different spatial and temporal resolutions. Thus, it is not possible with the conventional structure to construct a filter that captures the energy of moving objects with dissimilar spatial and temporal frequency characteristics such as large, fast objects (i.e., objects with high temporal frequency and low spatial frequency content), or small, slow objects (low temporal frequency and high spatial frequency content). In order to correct this problem, this section first presents the theory behind an unconventional, "motion-oriented" multiresolution analysis that decouples the spatial and temporal decomposition processes of

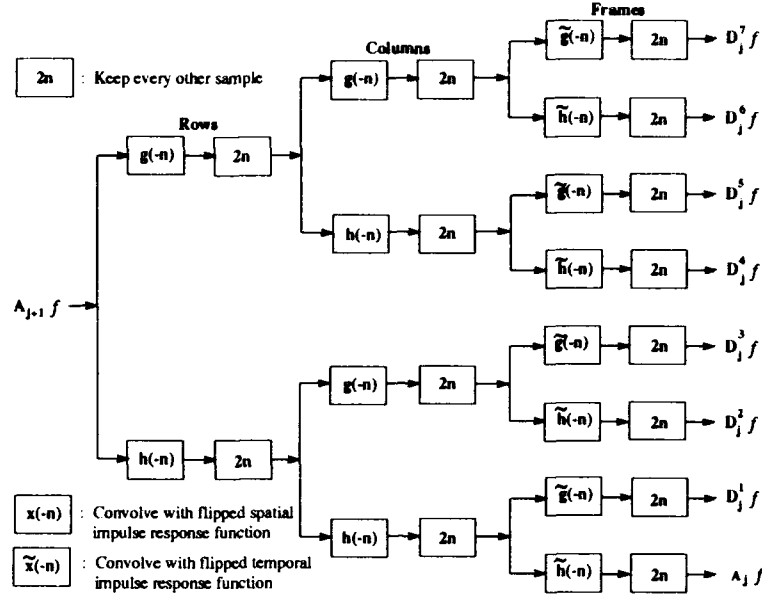


Figure 22. Oct-tree sub-band coding structure used to decompose the $j + 1$ st approximation coefficients into the j th approximation and detail coefficients in a conventional, non-homogeneous $L_2(\mathbb{R}^3)$ wavelet multiresolution analysis.

the conventional multiresolution analysis. A sub-band decomposition algorithm is then described which provides the ability to independently analyze spatial and temporal details in a 3D image sequence.

4.3.1 Decoupling the Spatial and Temporal Decomposition Process. The motion-oriented multiresolution wavelet analysis is based on the construction of an orthonormal basis for the “decoupled” spatio-temporal approximation space $V_{j,k}$. The definition of the decoupled spatio-temporal approximation space is given by

$$V_{j,k} = V_j^x \otimes V_j^y \otimes \tilde{V}_k^t = \overline{\text{Span}\{F(x, y, t) = f(x)g(y)h(t) \mid f \in V_j^x, g \in V_j^y \text{ and } h \in \tilde{V}_k^t\}} \quad (92)$$

where j represents spatial resolution, k represents temporal resolution and j is not necessarily equal to k . The corresponding orthonormal bases for $V_{j,k}$ are described in Theorem 5.

Theorem 5. For each $j \in \mathbb{Z}$ and $k \in \mathbb{Z}$, the set of functions $\{2^{j+\frac{k}{2}}\phi(2^j x - l)\phi(2^j y - m)\tilde{\phi}(2^k t - n) \mid (l, m, n) \in \mathbb{Z}^3\}$ forms an orthonormal basis for $V_{j,k}$.

Proof of Theorem 5. Let $\Phi_{j,k;(l,m,n)}(x,y,t) = 2^{j+\frac{k}{2}}\phi(2^jx-l)\phi(2^jy-m)\tilde{\phi}(2^kt-n)$. Then

$$\langle \Phi_{j,k;(l,m,n)}, \Phi_{j,k;(l',m',n')} \rangle = \begin{cases} 1 & \text{if } l = l' \text{ and } m = m' \text{ and } n = n' \\ 0 & \text{otherwise} \end{cases} \quad (93)$$

and the set of functions $\{\Phi_{j,k;(l,m,n)} \mid (l,m,n) \in \mathbb{Z}^3\}$ therefore forms an orthonormal set in $L_2(\mathbb{R}^3)$. Now, let F be a vector in $V_{j,k}$. By construction of $V_{j,k}$, $F(x,y,t) = f(x)g(y)h(t)$ for some $f \in V_j^x, g \in V_j^y$ and $h \in \tilde{V}_k^t$. Expressing f, g and h in terms of their respective orthonormal bases and rearranging terms yields

$$F = \sum_l \sum_m \sum_n 2^{j+\frac{k}{2}} \langle F, \Phi_{j,k;(l,m,n)} \rangle \Phi_{j,k;(l,m,n)} \quad (94)$$

Equation 94 implies F can be expressed as a Fourier series expansion of the orthonormal set $\{\Phi_{j,k;(l,m,n)} \mid (l,m,n) \in \mathbb{Z}^3\}$. Therefore, the Fourier Series Theorem ensures $\{\Phi_{j,k;(l,m,n)} \mid (l,m,n) \in \mathbb{Z}^3\}$ is an orthonormal basis for $V_{j,k}$. **Q.E.D.**

Now let $W_{j,k}$ represent the orthogonal complement of $V_{j,k}$ in $V_{j+1,k}$ such that

$$V_{j+1,k} = V_{j,k} \oplus W_{j,k} \quad (95)$$

Then, Theorem 6 describes an orthonormal basis for the spatial detail space $W_{j,k}$.

Theorem 6. Let ψ and $\tilde{\psi}$ be the one-dimensional wavelets generated by the scaling functions ϕ and $\tilde{\phi}$ respectively and let $W_{j,k}$ represent the orthogonal complement of $V_{j,k}$ in $V_{j+1,k}$. Then the three functions

$$\begin{aligned} \Psi_{j,k}^1(x,y,t) &= 2^{j+\frac{k}{2}}\phi(2^jx)\psi(2^jy)\tilde{\phi}(2^kt) \\ \Psi_{j,k}^2(x,y,t) &= 2^{j+\frac{k}{2}}\psi(2^jx)\phi(2^jy)\tilde{\phi}(2^kt) \\ \Psi_{j,k}^3(x,y,t) &= 2^{j+\frac{k}{2}}\psi(2^jx)\psi(2^jy)\tilde{\phi}(2^kt) \end{aligned} \quad (96)$$

are such that for each $(j,k) \in \mathbb{Z}^2$, $\{\Psi_{j,k}^p(x-l, y-m, t-n) \mid (l,m,n) \in \mathbb{Z}^3; p = 1, 2, 3\}$ forms an orthonormal basis for $W_{j,k}$.

Proof of Theorem 6. Given $V_{j,k}$ and the three multiresolution approximations in $L_2(\mathbb{R})$,

$$\{V_j^x \mid j \in \mathbb{Z}\}, \{V_j^y \mid j \in \mathbb{Z}\}, \{\tilde{V}_k^t \mid k \in \mathbb{Z}\} \quad (97)$$

$V_{j+1,k}$ can be expressed as

$$\begin{aligned} V_{j+1,k} &= V_{j+1}^x \otimes V_{j+1}^y \otimes \tilde{V}_k^t \\ &= (W_j^x \oplus V_j^x) \otimes (W_j^y \oplus V_j^y) \otimes \tilde{V}_k^t \end{aligned} \quad (98)$$

The right hand side of Equation 98 can be rewritten as follows

$$\begin{aligned} \text{RHS} &= [W_j^x \otimes W_j^y \otimes \tilde{V}_k^t] \oplus [W_j^x \otimes V_j^y \otimes \tilde{V}_k^t] \\ &\oplus [V_j^x \otimes W_j^y \otimes \tilde{V}_k^t] \oplus [V_j^x \otimes V_j^y \otimes \tilde{V}_k^t] \end{aligned} \quad (99)$$

Since $V_{j,k} = V_j^x \otimes V_j^y \otimes \tilde{V}_k^t$, the orthogonal complement, $W_{j,k}$, of $V_{j,k}$ in $V_{j+1,k}$ can be expressed as

$$W_{j,k} = V_{j+1,k} - V_{j,k} = [W_j^x \otimes W_j^y \otimes \tilde{V}_k^t] \oplus [W_j^x \otimes V_j^y \otimes \tilde{V}_k^t] \oplus [V_j^x \otimes W_j^y \otimes \tilde{V}_k^t] \quad (100)$$

The sets of functions $\{2^{\frac{j}{2}}\phi(2^j x - l) \mid l \in \mathbb{Z}\}$, $\{2^{\frac{j}{2}}\phi(2^j y - m) \mid m \in \mathbb{Z}\}$, and $\{2^{\frac{k}{2}}\tilde{\phi}(2^k t - n) \mid n \in \mathbb{Z}\}$ form orthonormal bases respectively for the $L_2(\mathbb{R})$ approximation spaces V_j^x , V_j^y and $\tilde{V}_k^t$. Additionally, the functions $\{2^{\frac{j}{2}}\psi(2^j x - l) \mid l \in \mathbb{Z}\}$ and $\{2^{\frac{j}{2}}\psi(2^j y - m) \mid m \in \mathbb{Z}\}$ form orthonormal bases respectively for the complementary spaces W_j^x and W_j^y . Thus, the set of functions $\{\Psi_{j,k}^p(x - l, y - m, t - n) \mid (l, m, n) \in \mathbb{Z}^3; p = 1, 2, 3\}$ forms an orthonormal basis for $W_{j,k}$. **Q.E.D.**

A straightforward consequence of Theorem 5 is that $V_{j,k}$ is contained in $V_{j',k'}$ if and only if $j \leq j'$ and $k \leq k'$. This fact is illustrated by the lattice of spaces shown in Figure 23. Here, the chain of spaces comprising the “conventional,” non-homogeneous 3D multiresolution analysis lies along the diagonal formed when $j = k$. The remaining subspaces are created by independently decomposing the conventional spaces along spatial (vertical) and temporal (horizontal) lines. In this illustration a finite decomposition beginning at $V_{2,2}$ is assumed. The detail spaces highlighted by gray squares are obtained by vertically decomposing the originally sampled signal contained in $V_{2,2}$ using the algorithm

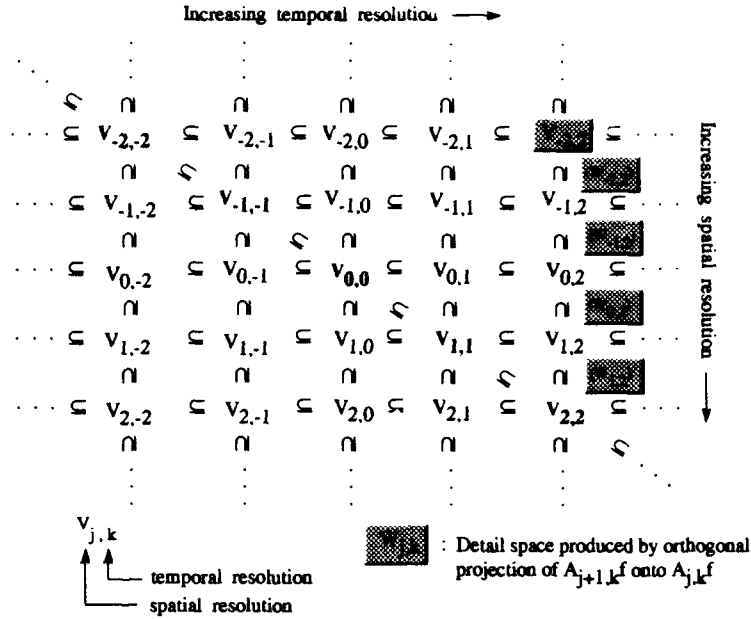


Figure 23. Array of embedded subspaces formed by further decomposing conventional multiresolution spaces (represented by diagonal $k = j$) along vertical (spatial) and horizontal (temporal) lines.

described below. These spaces capture the spatial details in the image sequence for a fixed resolution in time of $k = 2$. Theorem 6 ensures the highlighted detail spaces are orthogonal. Theorem 7 now shows that each detail space, $W_{j,k}$, can be decomposed in time to produce orthogonal, temporal detail spaces for a fixed resolution in space. The temporal decomposition is based on a special case of Coifman and Meyer's wavelet packet theory as proved by I. Daubechies (12, 14).

Theorem 7. Let $W_{j,k}^p$ ($p = 1, 2, 3$) represent the space spanned by integer translations in space and time of the function $\Psi_{j,k}^p(x, y, t) = \Psi_j^p(x, y)2^{\frac{k}{2}}\tilde{\phi}(2^k t)$ where,

$$\begin{aligned}
 \Psi_j^1(x, y) &= 2^j \phi(2^j x) \psi(2^j y) \\
 \Psi_j^2(x, y) &= 2^j \psi(2^j x) \phi(2^j y) \\
 \Psi_j^3(x, y) &= 2^j \psi(2^j x) \psi(2^j y)
 \end{aligned} \tag{101}$$

Define the functions,

$$\begin{aligned}\Psi_{j,k}^{p1}(x, y, t) &= \Psi_j^p(x, y) \sum_n 2^{\frac{k}{2}} \tilde{h}_n \tilde{\phi}(2^k t - n) \\ \Psi_{j,k}^{p2}(x, y, t) &= \Psi_j^p(x, y) \sum_n 2^{\frac{k}{2}} \tilde{g}_n \tilde{\phi}(2^k t - n)\end{aligned}\quad (102)$$

then $\{\Psi_{j,k}^{p1}(x-l, y-m, t-2n), \Psi_{j,k}^{p2}(x-l, y-m, t-2n) \mid (l, m, n) \in \mathbb{Z}^3\}$ form an orthonormal basis for

$$W_{j,k}^p = \overline{\text{Span}\{\Psi_{j,k}^p(x-l, y-m, t-n) \mid (l, m, n) \in \mathbb{Z}^3\}}$$

Proof. Consider the following Lemma which describes a special case of Coifman and Meyer's wavelet packet theory as proved by I. Daubechies (12, 14).

Lemma 1. Let f be any function such that the $f(t-n)$, $n \in \mathbb{Z}$, are orthonormal. Define the functions

$$\begin{aligned}F^1(t) &= \sum_n h_n f(t-n) \\ F^2(t) &= \sum_n g_n f(t-n)\end{aligned}\quad (103)$$

Then $\{F^1(t-2m), F^2(t-2m) \mid m \in \mathbb{Z}\}$ forms an orthonormal basis for $\overline{\text{Span}\{f(t-n) \mid n \in \mathbb{Z}\}}$.

Since the functions $2^{\frac{k}{2}} \tilde{\phi}(2^k t - n)$, $n \in \mathbb{Z}$, are orthonormal, Lemma 1 implies $\{F_k^1(t-2m), F_k^2(t-2m) \mid m \in \mathbb{Z}\}$ forms an orthonormal basis for $\overline{\text{Span}\{\tilde{\phi}(2^k t - n) \mid n \in \mathbb{Z}\}}$ where

$$\begin{aligned}F_k^1(t) &= \sum_n h_n \tilde{\phi}(2^k t - n) \\ F_k^2(t) &= \sum_n g_n \tilde{\phi}(2^k t - n)\end{aligned}\quad (104)$$

Now, let $\Psi_{j,k}^p(x-l_o, y-m_o, t) = \Psi_j^p(x-l_o, y-m_o) 2^{\frac{k}{2}} \tilde{\phi}(2^k t)$ where the integer pair $(l_o, m_o) \in \mathbb{Z}^2$ is chosen arbitrarily. Next, define the functions $\Psi_{j,k}^{p1}$ and $\Psi_{j,k}^{p2}$ as follows:

$$\begin{aligned}\Psi_{j,k}^{p1}(x-l_o, y-m_o, t) &= \Psi_j^p(x-l_o, y-m_o) F_k^1(t) \\ \Psi_{j,k}^{p2}(x-l_o, y-m_o, t) &= \Psi_j^p(x-l_o, y-m_o) F_k^2(t)\end{aligned}\quad (105)$$

Then, the set of functions $\{\Psi_{j,k}^{p1}(x-l_o, y-m_o, t-2n), \Psi_{j,k}^{p2}(x-l_o, y-m_o, t-2n) \mid n \in \mathbb{Z}\}$ forms an orthonormal basis for $\text{Span}\{\Psi_{j,k}^p(x-l_o, y-m_o, t-n) \mid n \in \mathbb{Z}\}$. But (l_o, m_o) chosen arbitrarily implies $\{\Psi_{j,k}^{p1}(x-l, y-m, t-2n), \Psi_{j,k}^{p2}(x-l, y-m, t-2n) \mid (l, m, n) \in \mathbb{Z}^3\}$ form an orthonormal basis for $\text{Span}\{\Psi_{j,k}^p(x-l, y-m, t-n) \mid (l, m, n) \in \mathbb{Z}^3\}$. **Q.E.D.**

4.3.2 Discrete, Motion-Oriented Decomposition Algorithm. The previous section ensures the spatial and temporal decomposition processes in the conventional multiresolution analysis can be decoupled to generate multiple temporal resolutions of a 3D signal for a fixed spatial resolution. Furthermore, the spaces containing the temporal detail signals are orthogonal across all spatial and temporal resolution levels. This section describes an $\mathcal{O}(N^3)$ sub-band decomposition algorithm that produces the coefficients obtained by orthogonally projecting a 3D signal onto each of these orthogonal detail spaces. It is further shown that the filter bank produced by this unconventional decomposition algorithm yields a set of independent spatial-temporal channels for locating vertical edges, horizontal edges and corners of objects moving at different speeds.

In order to describe the algorithm, consider the problem of analyzing the motion of a two dimensional object traveling in an $N \times N \times N$ image sequence. In this case, $V_{j,k}$ represents the closed linear space formed by the tensor product $V_j^x \otimes V_j^y \otimes V_k^t$. Additionally, let $A_{0,0}f$ (i.e., the discrete projection of the original signal onto the space $V_{0,0}$) represent the sampled 3D image sequence. Finally, let $D_{-1,0}^p f$ represent the discrete projection of the signal onto the spatial detail spaces $W_{-1,0}^p$ where it is understood that $p = 1, 2, 3$. A visualization of the decomposition process is shown in Figure 24.

In the first stage of the decomposition algorithm, $A_{0,0}f$ is decomposed *spatially* into the approximation and detail signals $A_{-1,0}f$ and $D_{-1,0}^p f$ respectively by convolving the rows and columns of each frame in $A_{0,0}f$ with flipped versions of the spatial filters h and g , and decimating the spatial dimensions by a factor of two. This process is illustrated for arbitrary spatial and temporal resolution levels j and k in Figure 25. The spatial algorithm is then applied recursively to each subsequent spatial approximation signal, $A_{-j,0}$; $j = 1, 2, 3, \dots$, to generate a sequence of signals which captures the spatial details between successively smaller spatial resolutions for the temporal resolution $k = 0$. The spatial approximation signals, $A_{-j,0}$; $j = 1, 2, 3, \dots$, produced by this process are represented by the lightly shaded planes in Figure 24. The darker planes represent the spatial detail signals $D_{-j,0}^p f$; $p = 1, 2, 3$; $j = 1, 2, 3, \dots$

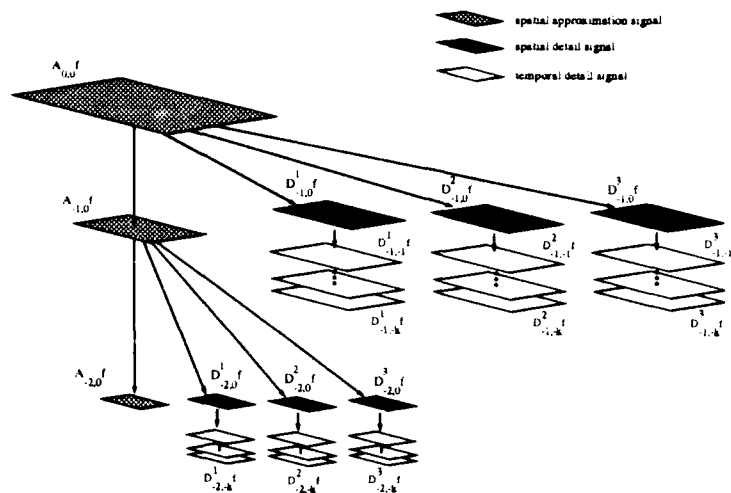


Figure 24. A visualization of the 3D motion-oriented wavelet decomposition process.

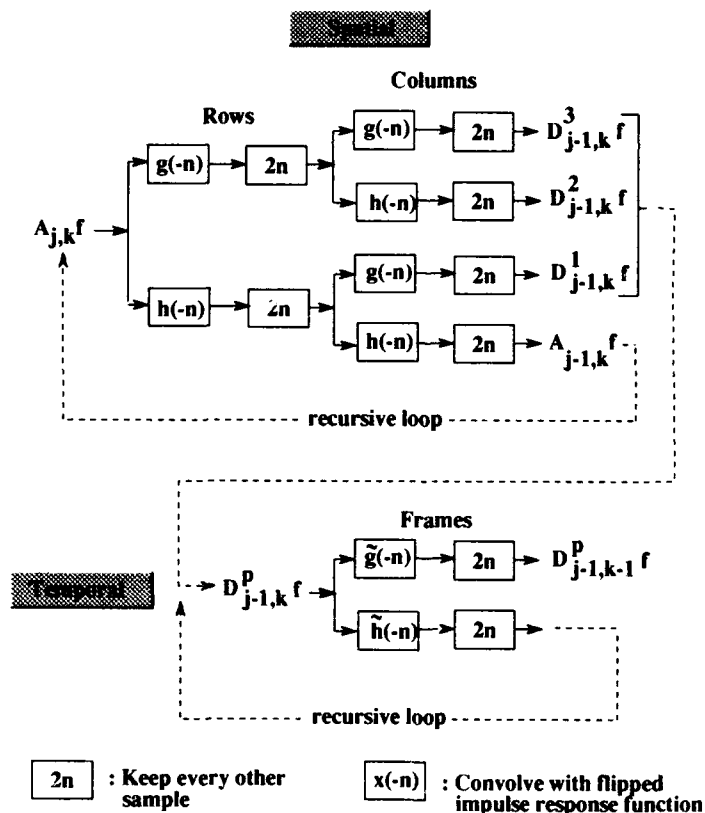


Figure 25. Spatial, and temporal decomposition algorithms for 3D motion-oriented multiresolution wavelet analysis. Decomposition is shown for arbitrary spatial and temporal resolutions levels j and k .

Assuming the number of coefficients in the wavelet filter is small compared to the number of samples, N , in each dimension of the image sequence, the computational complexity of this stage of the algorithm is found by determining the total number of values computed by the spatial decomposition process. To this end, first note that the number of samples at each spatial decomposition level are half the number at the next higher level. Thus, if the spatial dimensions of a frame in the originally sampled signal are $N \times N$, then the dimensions of a frame at the next lower spatial decomposition level are $\frac{N}{2} \times \frac{N}{2}$. Furthermore, since four signals are produced by the spatial decomposition process (1 approximation and 3 detail), the total number of values computed by the first spatial decomposition is $\frac{N^2}{4} + \frac{N^2}{4} + \frac{N^2}{4} + \frac{N^2}{4} = N^2$. Continuing the process, the next spatial decomposition produces a total of $\frac{N^2}{4}$ values per frame, and so on. Letting the number of spatial decompositions go to infinity then gives an upper bound on the number of spatial values computed *per frame* of $\frac{4N^2}{3}$. Finally, assuming there are N frames in the image sequence, the total number of spatial values computed in the spatial decomposition stage of the algorithm is then $N \cdot \frac{4N^2}{3} = \frac{4N^3}{3}$.

In the next stage of the algorithm, the first level spatial detail signals $D_{-1,0}^p f$; $p = 1, 2, 3$ are decomposed in time by convolving flipped versions of the temporal filters $\tilde{h}$ and $\tilde{g}$ across all frames at each spatial location and decimating the temporal dimension by a factor of two (Figure 25). The temporal decomposition algorithm is then applied in a cascade fashion to each of the temporal approximation signals to yield a set of temporal detail signals, $D_{-1,k}^p f$; $p = 1, 2, 3$; $k = 1, 2, 3, \dots$, for each spatial detail signal in the first spatial decomposition level. This process is then repeated for each spatial detail signal $D_{-j,0}^1 f$, $D_{-j,0}^2 f$, and $D_{-j,0}^3 f$; $j = 2, 3, 4, \dots$ at each stage of the spatial decomposition process. The temporal detail signals produced by this process are represented by the unshaded planes in Figure 24.

In order to determine the computational complexity of the temporal decomposition stage of the algorithm, note that the upper bound on the number of temporal values computed over all temporal decomposition levels at one spatial location is $2N$. Consequently, given that the number of spatial locations produced in the spatial decomposition process is bounded by $\frac{4N^2}{3}$, the total number of values computed in the temporal stage of the algorithm is then $2N \cdot \frac{4N^2}{3} = \frac{8N^3}{3}$. Finally, adding the upper bounds on the spatial and temporal decomposition processes yields an upper bound of $\frac{8N^3}{3} + \frac{4N^3}{3} = 4N^3$ for the total number of values computed in the spatio-temporal decomposition

process. Thus, the computational complexity of the discrete motion-oriented multiresolution wavelet decomposition algorithm is $\mathcal{O}(N^3)$.

By repeatedly decomposing the temporal information contained in each spatial detail signal one gains the ability to independently zoom-in and zoom-out on spatial and temporal details in the scene. For example, assuming the size of a moving object corresponds to the spatial resolution $j = -3$, its speed can be approximated by comparing the magnitude of the coefficients contained in the temporal detail signals $D_{-3,k}^1 f$, $D_{-3,k}^2 f$, and $D_{-3,k}^3 f$; $k = 1, 2, 3, \dots$ (recall that in the conventional decomposition scheme, the analysis is restricted to temporal detail information contained in the space $W_{-3,-3}$). Furthermore, Theorems 6 and 7 guarantee that the detail spaces generated by the spatial and temporal decomposition processes in the motion-oriented algorithm are orthogonal. Therefore, the main lobes of the spatio-temporal frequency spectrums of the basis functions associated with these spaces have essentially non-overlapping regions of support in the Fourier frequency domain. This behavior is illustrated in Figure 26.

Figure 26 shows the supporting regions in the positive half of the temporal frequency plane of the 3D multiresolution motion analysis filters. Notice, that the filter's passbands in the 2D spatial frequency plane ($f_t = 0$) are identical to those produced by the conventional 3D multiresolution decomposition algorithm described in Section 3.1. However, unlike the frequency spectrum generated by the conventional $L_2(\mathbb{R}^3)$ wavelet multiresolution analysis (Chapter III), which contains only low pass and band pass support regions for each of the horizontal, vertical and diagonal spatial detail filters in the spatial frequency plane, the new frequency spectrum contains a bank of temporal frequency bandpass filters for each spatial orientation. Viewed from a motion analysis perspective, this unique and unconventional filter bank now provides the flexibility to discriminate objects moving in a 3D image sequence with dissimilar spatial and temporal frequency characteristics (e.g., small objects traveling slow and large objects traveling fast).

In the discussions that follow, it will sometimes be easier to first explain a particular concept as it applies to the problem of computing the velocity of a 1D object moving in a 2D spatio-temporal image sequence. Thus, as an aid to the reader, Figure 27 shows the spatio-temporal detail spaces obtained by applying the motion-oriented decomposition algorithm to a 2D image sequence. The shaded regions in Figure 27 each represent the frequency support of one detail signal filter generated by the 2D motion wavelet decomposition process. The narrow, white vertical strip in the center represents the filter

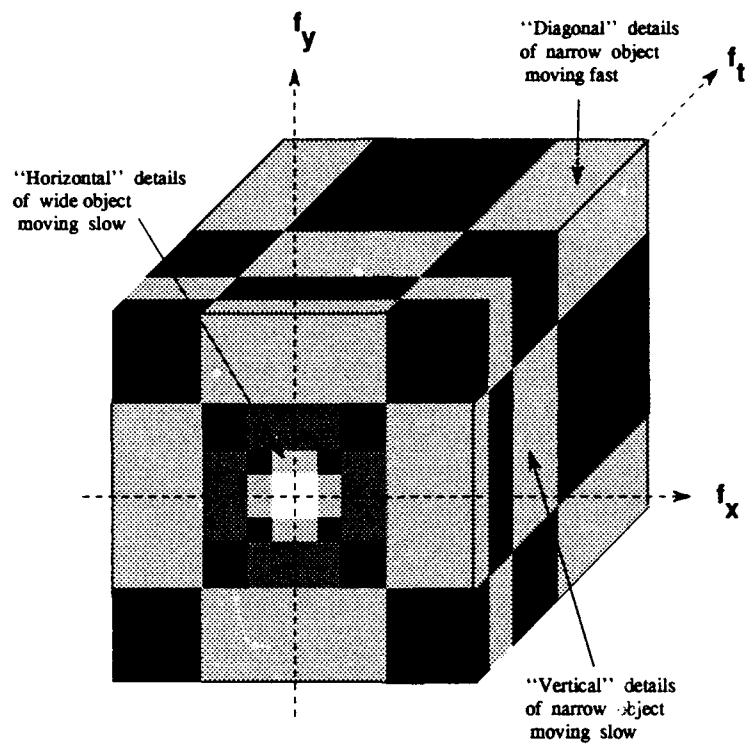


Figure 26. A visualization of the frequency support in the Fourier plane of the basis functions for each space generated by the 3D wavelet multiresolution motion decomposition.

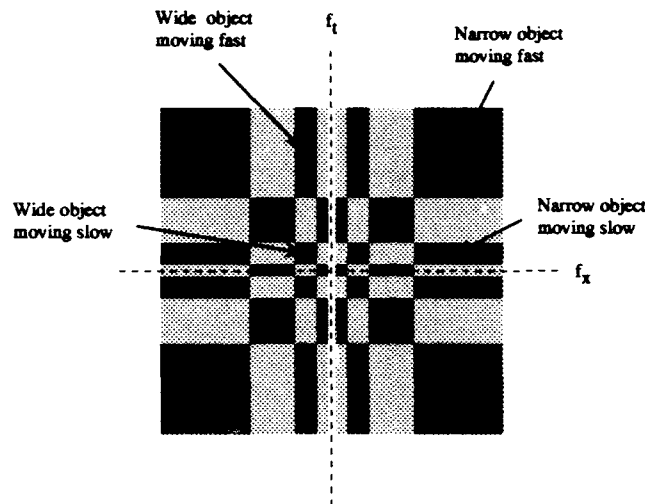


Figure 27. A visualization of the frequency support in the Fourier plane of the basis functions for each space generated by the 2D wavelet multiresolution motion decomposition.

associated with the last approximation space generated by the spatial decomposition algorithm. Just as in Mallat's conventional 1D multiresolution analysis, the abscissa of the frequency plane divides the original signal into fine and coarse spatial detail signals. The outermost lighter region contains the spatial details of narrow objects (i.e., high spatial frequencies), while the innermost dark region captures the spatial details of wide objects. The ordinate axis divides each spatial detail signal into multiple temporal detail signals that capture temporal frequency components associated with multiple speeds. Thus, for example, the large dark square in the upper right hand corner captures narrow, *fast* moving 1D objects; while the dark region two temporal resolution levels beneath it captures narrow, *slow* moving 1D objects.

The filter banks produced by the 2D and 3D motion-oriented wavelet decomposition algorithm yield a set of independent spatio-temporal channels for locating vertical edges, horizontal edges and corners of objects moving at different speeds. The motion-oriented filter bank was generated using a rapid sub-band coding scheme in which a discretely sampled input signal was decomposed independently in space and time. By discretely sampling the input signal, one risks the possibility of spatial and/or temporal aliasing. Spatial aliasing is a common image processing problem that is typically handled by a 2D lowpass filtering operation (16, 20). The temporal aliasing problem, particularly as it pertains to 2D objects moving in a 3D image sequence, is less commonly discussed in the literature.

Therefore, the following section examines the effect of spatio-temporal aliasing on Fourier frequency motion analysis.

4.4 Spatio-Temporal Aliasing and Fourier Frequency Motion Analysis

In Chapter II, a simple relationship was derived between the speed of a sinusoidal grating and its temporal frequency. Additionally, it was shown that the temporal frequencies of a more complex object moving at a constant velocity are related to the object's velocity components v_x and v_y . In both cases, the temporal frequency bandwidth increased proportionately with the velocity of the object. The velocity of an object therefore plays a critical role in determining the temporal sampling rates required to prevent aliasing in a discretely sampled signal. For example, consider the case of the simple traveling sinusoid

$$f(x, y, t) = \cos(ax + by + ct) \quad (106)$$

It was shown in Section 2.4 that the x and y components of the sinusoid's velocity vector V are given by

$$v_x = -\frac{ac}{a^2 + b^2} \quad (107)$$

$$v_y = -\frac{bc}{a^2 + b^2} \quad (108)$$

If the spatial sampling frequency exceeds the Nyquist limit, and if the temporal sampling frequency is given by C , then aliasing will not occur provided the temporal frequency $c \leq \frac{C}{2}$. This implies that, for a fixed spatial frequency, the magnitude of the velocity vector $\|V\|$ must be such that

$$\begin{aligned} \|V\| &= \sqrt{v_x^2 + v_y^2} \\ &= \sqrt{\left(\frac{ac}{a^2 + b^2}\right)^2 + \left(\frac{bc}{a^2 + b^2}\right)^2} \\ &\leq \frac{C}{2\sqrt{a^2 + b^2}} \end{aligned} \quad (109)$$

The relationship in Equation 109 shows that the magnitude of the velocity is inversely proportional to the magnitude of the sinusoid's spatial frequency. Consequently, as the spatial frequency of the sinusoid decreases, larger velocities are allowed before temporal aliasing occurs. Of course, this argument

assumes the moving object is a sinusoid of a given spatial frequency. And since a sinusoid always travels perpendicular to its brightness contour, it was only necessary to consider the magnitude of the velocity vector. In reality, however, the aliasing limits of more complicated objects are determined by both the magnitude and the direction of their velocity. This topic is discussed next.

Recall from Chapter II that the Fourier transform of an object moving with the constant velocity components (v_x, v_y) is given by

$$\mathcal{F}\{f(x - v_x t, y - v_y t)\} = F(f_x, f_y, f_t + v_x f_x + v_y f_y) \quad (110)$$

which implies the 2D Fourier transform of the moving object is shifted onto the plane given by

$$f_t = -(v_x f_x + v_y f_y) \quad (111)$$

If the temporal sampling frequency is given by F_t , and if one again assumes that the spatial sampling frequency exceeds the Nyquist limit, then temporal aliasing will not occur provided

$$\begin{aligned} f_t &= v_x f_x + v_y f_y \\ &= ||V|| ||f|| \cos(\phi_V - \phi_f) \\ &\leq \frac{F_t}{2} \end{aligned} \quad (112)$$

where $||f||$ is the magnitude of the spatial frequency pair (f_x, f_y) , $||V||$ is the magnitude of the velocity vector, ϕ_f is the angle of the spatial frequency pair, ϕ_V is the angle of the velocity vector and the minus sign in Equation 111 has been neglected. Equation 112 shows that the presence or absence of temporal aliasing depends on a vector product relationship between the spatial frequency content of the object, its speed *and* its direction of motion in the sense that the temporal frequency depends on the cosine of the angle between the direction of motion and the direction of a particular spatial frequency. The following paragraphs provide examples of temporal aliasing in the frequency representation of an image sequence that contains a single moving object.

Consider the 2D gaussian moving along a 45° trajectory as shown in Figure 28a). The image volume is $64 \times 64 \times 64$. The Fourier transform of the moving object lies along the plane in Figure 28b), where the “slope” of the plane (i.e., the tangent of the angle between the f_t axis and the nearest

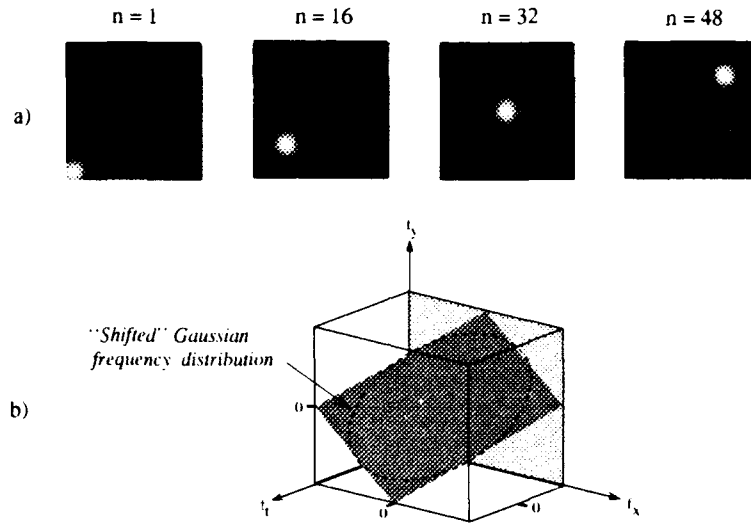


Figure 28. a) A gaussian moving along a 45° trajectory. b) The plane in the Fourier domain contains the frequency components of the moving object. The shaded side of the frequency volume represents the spatial frequency components associated with the most negative temporal frequency component.

vector contained in the plane) is determined by the speed of the gaussian. The lightly shaded surface of the frequency volume represents the spatial frequency components associated with the most negative temporal digital frequency component $f_t = -\pi$.

Assume now that the velocity components of the moving gaussian are 1 frame/sec. in the x and y directions. Also assume that input signal is sampled so that the spatial and temporal sampling rates do not violate the Nyquist sampling criteria, and that the cutoff radius of the circularly symmetric DFT of the gaussian is approximately $\frac{\pi}{\sqrt{2}}$. The spatial frequency components in the $f_t = -\pi$ frequency plane will then form a single line as shown by the density plot of the moving object's FFT contained in Figure 29a). Furthermore, Equation 112 implies the maximum temporal frequency of the object, $f_{t_{max}}$, occurs at the spatial frequency that lies in the direction of motion (for the circularly symmetric Gaussian frequency distribution). The spatial frequency coordinates at which this occurs are $(\frac{\pi}{2}, \frac{\pi}{2})$, yielding a maximum temporal frequency of

$$\begin{aligned}
 f_{t_{max}} &= 1 \frac{\text{frame}}{\text{sec}} \cdot \frac{\pi \text{ cycles}}{2 \text{ frame}} + 1 \frac{\text{frame}}{\text{sec}} \cdot \frac{\pi \text{ cycles}}{2 \text{ frame}} \\
 &= \pi \frac{\text{cycles}}{\text{sec}}
 \end{aligned} \tag{113}$$

which is equal to the digital Nyquist cutoff frequency. Now consider what occurs when the variance of the gaussian is reduced so that the spatial frequency radius is increased to approximately $\sqrt{2}\pi$ radians. If the object's x and y velocity components remain constant at 1 frame/sec., the maximum digital temporal frequency now becomes

$$\begin{aligned} f_{t,max} &= 1 \frac{\text{frame}}{\text{sec}} \cdot \pi \frac{\text{cycles}}{\text{frame}} + 1 \frac{\text{frame}}{\text{sec}} \cdot \pi \frac{\text{cycles}}{\text{frame}} \\ &= 2\pi \frac{\text{cycles}}{\text{sec}} \end{aligned} \quad (114)$$

which is twice the digital Nyquist limit. Thus, one would expect to see aliased frequency components near the temporal frequency borders of the FFT frequency volume. This is clearly the case as shown by the $f_t = -\pi$ plane contained in Figure 29b). Here, the aliased components appear as a second line in the lower left corner of the frequency plane. Also, the length of the line is greater than in a) since the frequency cutoff radius has doubled.

Equation 112 also implies that temporal aliasing will occur when large objects (with small spatial frequency magnitudes) travel too fast. This is demonstrated by the double lines in Figure 29c). Here, the velocity of the gaussian is 2 frames/sec. in both directions and the maximum spatial frequency magnitude has been reduced to its previous value of $\frac{\pi}{\sqrt{2}}$, yielding a maximum temporal frequency of

$$\begin{aligned} f_{t,max} &= 2 \frac{\text{frame}}{\text{sec}} \cdot \frac{\pi \text{ cycles}}{2 \text{ frame}} + 2 \frac{\text{frame}}{\text{sec}} \cdot \frac{\pi \text{ cycles}}{2 \text{ frame}} \\ &= 2\pi \frac{\text{cycles}}{\text{sec}} \end{aligned} \quad (115)$$

The maximum spatial frequency magnitude is once again the same size as the first example, therefore the line of spatial frequency components in the density plot is shorter than those in b). Also, because the object is traveling faster than the objects in the other two examples, the slope of the plane decreases (i.e., the plane lies closer to the f_t axis) and the lines in c) lie closer to each other.

In order to prevent temporal aliasing, Equation 112 implies one can spatially filter each frame in the image sequence to limit the magnitude of the spatial frequencies. For example, a circular filter with a radius of $\sqrt{f_{x0}^2 + f_{y0}^2}$ might be a good choice where f_{x0} and f_{y0} are the cutoff frequencies of the filter. One might then assume a worst case scenario in which the direction of motion would lie in the direction of the spatial frequency with the largest magnitude contained within the passband of the filter.

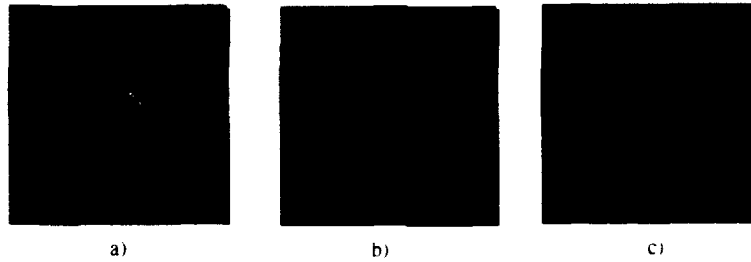


Figure 29. a) FFT of a single line of spatial frequency components in the $f_t = -\pi$ temporal frequency plane generated by a moving gaussian whose temporal frequencies satisfy the inequality in Equation 112. b) Two lines appear in the $f_t = -\pi$ plane as a result of temporal aliasing. The object travels with the same velocity as a), but the object size decreases by a factor of two. c) Temporal aliasing again forms two lines in the $f_t = -\pi$ plane. Now, however, the object size is the same as in a), but the velocity components have both doubled.

This assumption sets the cosine term equal to one and reduces Equation 112 to Equation 109. Under these circumstances, one can then choose the appropriate temporal sampling frequency to ensure the maximum expected speed of any object moving in the scene does not violate the inequality in Equation 109.

4.5 Applications and Results

The purpose of this section is to demonstrate the capabilities and the limitations of the motion-oriented wavelet multiresolution analysis by applying it to several different image sequences. The first two tests are designed to show that the motion-oriented decomposition algorithm, unlike the conventional $L_2(\mathbb{R}^3)$ algorithm in Chapter III, can simultaneously look across different scales in space and time to differentiate between 1) two equally sized objects traveling at different speeds, and 2) two different sized objects traveling at different speeds. In the third test, the motion-oriented motion algorithm is applied to real IR imagery of a tank moving across open terrain. The outcome of this test demonstrates the algorithm's ability to zoom-in and zoom-out on spatial and temporal details in a noisy scene. The results are also briefly compared with the output of a simple frame-differencing motion segmentation technique. Finally, the algorithm is applied to a synthetic image sequence containing two equally-sized objects traveling at the same speed but in opposite directions. This test demonstrates a fundamental limitation of the motion-oriented wavelet decomposition algorithm - it is not directionally selective. A solution to this problem is presented in Chapter V. Each of the tests conducted in the

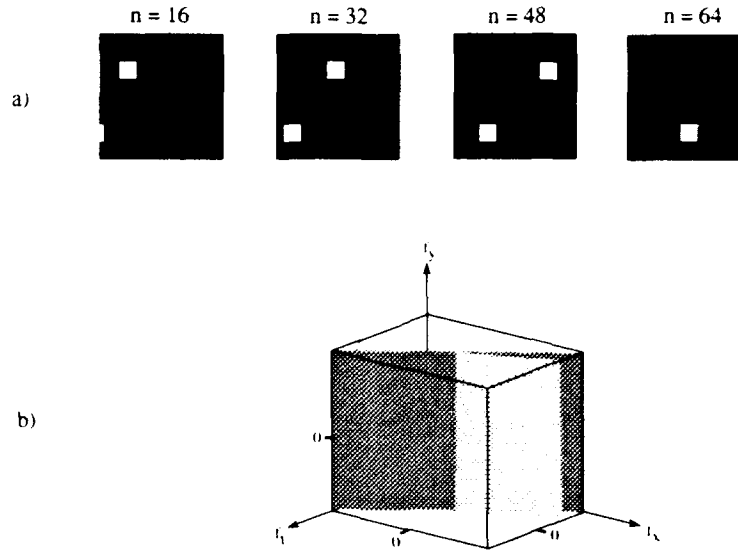


Figure 30. a) Several frames of 64×64 synthetic, grayscale imagery containing two equally sized rectangles traveling at different speeds. The speed of the upper rectangle is twice that of the lower rectangle. n represents a frame in the image sequence. b) A visualization of the planes containing the Fourier transforms of both rectangles. The darker plane corresponds to the faster rectangle.

chapter employed a Daubechies 4 QMF pair for spatial decomposition and a Daubechies 12 QMF pair for temporal decomposition (14). This yields greater resolution along the temporal frequency axis with which to separate the speeds of the moving objects. Since the objects are identical, spatial resolution is less important, allowing for the use of a more computationally efficient 4 tap spatial filter.

The first sequence of images consists of 128 frames of 64×64 synthetic, grayscale imagery. The image sequence, shown in Figure 30a), contains two equally sized rectangles traveling horizontally across an image plane at two different speeds. The upper object travels at one frame per second, and the lower object travels at one-half frame per second. Since the vertical velocity, v_y , of both rectangles equals zero, their Fourier transforms will lie on the planes given by the equation $f_t = -f_x v_x$ where v_x is either one or one-half frames per second. If the largest digital spatial frequency of both objects is π , then the planes will appear as shown in Figure 30b).

Now consider the horizontal plane of the Fourier transform taken through the largest positive digital spatial frequency $f_y = \pi$ as shown in Figure 31. The frequency support of the wavelet filters generated by several decompositions in space (f_x) and time are overlayed on this plane. Considering

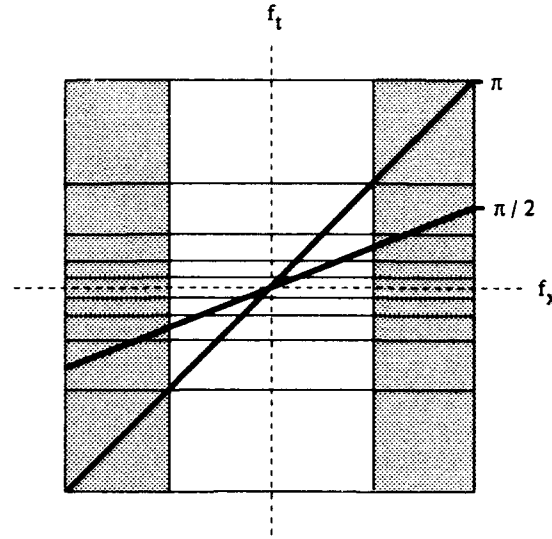


Figure 31. Frequency supports of the wavelet filters generated by several decompositions in time for a plane taken through the FFT of the image sequence in Figure 30a) at the spatial frequency $f_y = \pi$. The dark lines represent the 2D projections of the Fourier transforms of the moving objects. The spatial frequency axis f_y points out of the paper.

only the spatial frequencies surrounding $f_x = \pi$, the filters generated at each step in the temporal decomposition process are highlighted in gray. Note that the Fourier transform of the fastest object intersects the filter with digital center frequencies $f_x = \pi, f_y = \pi, f_t = \pi$, while the Fourier transform of the slower object lies through the filter located at $f_x = \pi, f_y = \pi, f_t = \frac{\pi}{2}$. Further note that although the center frequencies are specified here by their positive spatial and temporal frequencies, the filters are actually symmetric around all three axes (recall Figure 26).

In order to segment the two horizontally moving objects, the representations in Figures 30 and 31 suggest choosing the wavelet coefficients associated with the first spatial decomposition level and either the first or the second temporal decomposition levels. Furthermore, at either temporal level, one can also choose between filters that extract diagonal or vertical object features. Figure 32 shows both cases for the first and second temporal decomposition levels. Here, the outputs were thresholded to eliminate the small amount of energy captured in overlapping frequency bands of neighboring filters. Incidentally, since the planes do not pass through the wavelet filters associated with horizontal features, it is not possible to segment horizontal features. This is consistent with the aperture problem described

in Chapter II which prevents the measurement of motion in a direction parallel to a brightness contour (in this case, a horizontal line moving horizontally).

The second image sequence contains two differently sized objects traveling at two different speeds. Again, the purpose of this test is to demonstrate the algorithm's ability to extract information at different scales in space and time. Specifically, recall that the conventional wavelet multiresolution analysis restricts the analysis of motion to the same resolution in space and time. That is, the filters produced by this approach are tuned to either large, slow objects (low frequencies in space and time) or small, fast objects (high spatial and temporal frequencies). The conventional approach therefore cannot extract moving objects with dissimilar spatial and temporal frequency characteristics, such as large/fast or small/slow objects. This experiment shows the motion-oriented multiresolution analysis segments objects with both types of dissimilar 3D frequency spectrums. The $64 \times 64 \times 64$ grayscale image sequence is shown in Figure 33. The larger of the two rectangles is traveling vertically at two frames per second, while the smaller rectangle's speed is one frame per second. The dimensions of the large rectangle are twice those of the smaller rectangle.

In the second test set, the horizontal velocities of both objects equal zero, so their Fourier transforms lie on the planes given by $f_t = -f_y v_y$ where v_y is either one or two frames per second. If the largest digital spatial frequencies of the small and large objects are π and $\approx \frac{\pi}{2}$ respectively, their Fourier transforms will lie on the planes shown in Figure 33b). Following the previous example, consider the vertical plane of the Fourier transform taken through the largest positive digital temporal frequency $f_t = \pi$ as shown in Figure 34. The frequency supports of the wavelet filters for several spatial resolutions are overlayed on the $f_t = \pi$ plane, and the dark lines represent the intersections of the planar frequency supports of the two object with this plane.

Figure 34 suggests that the two objects can be segmented in frequency space by filtering the image sequence with the wavelet filters associated with either the light or dark gray regions of the $f_t = \pi$ plane. This corresponds to a wavelet decomposition of one resolution level in time and either one (dark gray) or two (light gray) resolution levels in space. Figure 35 shows the wavelet coefficients obtained by such a decomposition. Assuming the image sequence represents the coefficients associated with the projection of the signal onto the $j = 0, k = 0$ approximation space, then Figure 35a) contains the magnitude of the coefficients of the projection onto the $j = -1, k = -1$ resolution level, and the coefficients in Figure 35b) are from the $j = -2, k = -1$ resolution level. Two different detail

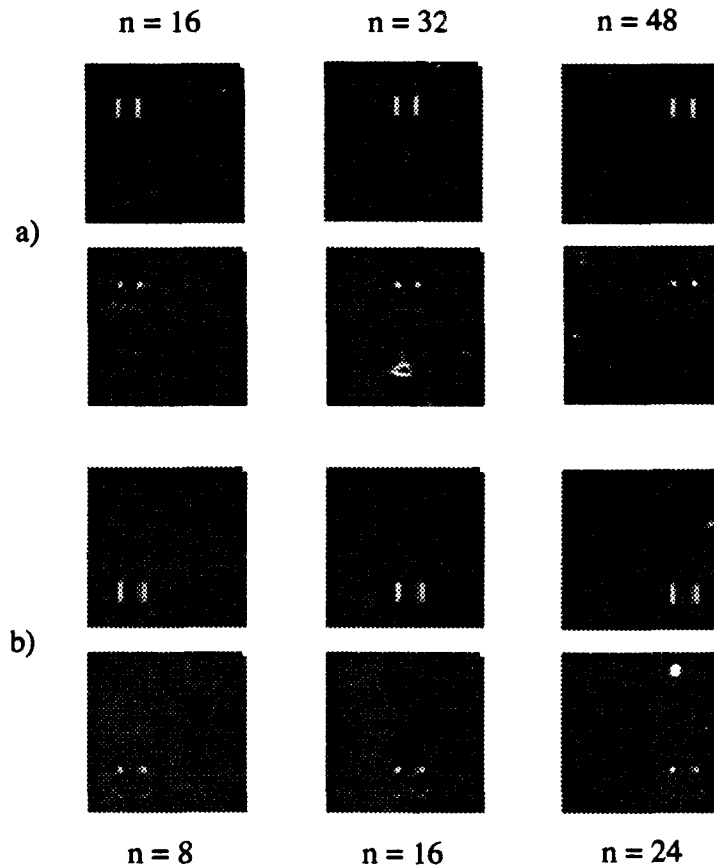


Figure 32. a) Segmenting the diagonal and vertical features of the faster object by decomposing the input signal one level in space and one level in time. The slower object is completely attenuated by the motion-oriented filter bank. The dimensions of the resulting coefficient sequence are $64 \times 64 \times 64$. b) Segmenting the slower object by decomposing one level in space and two levels in time. The coefficient sequence dimensions are $64 \times 64 \times 32$ (row, column, frame). Here n represents a frame in a coefficient sequence. In this case, the faster object is completely eliminated by the filter bank.

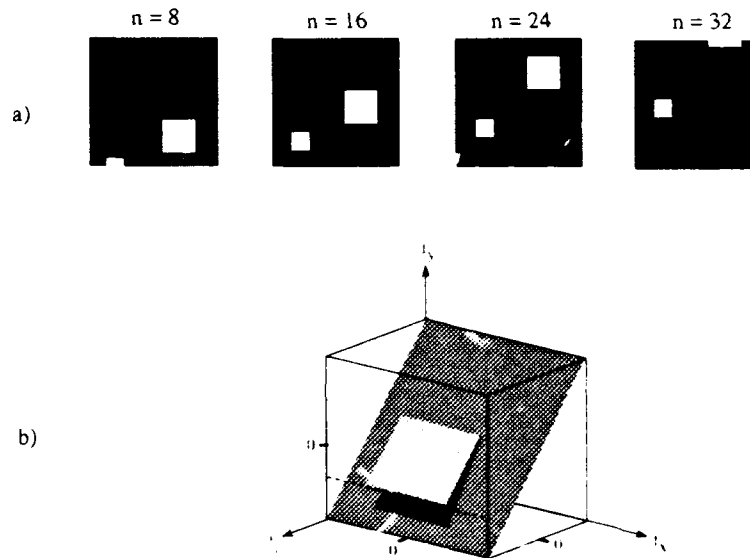


Figure 33. a) Several frames of 64×64 synthetic, grayscale imagery containing two differently sized rectangles traveling at two different speeds. The larger rectangle is traveling at twice the velocity as the small rectangle. b) A visualization of the planes containing the Fourier transforms of both rectangles. The Fourier transform of the larger rectangle lies on the narrower, lighter plane.

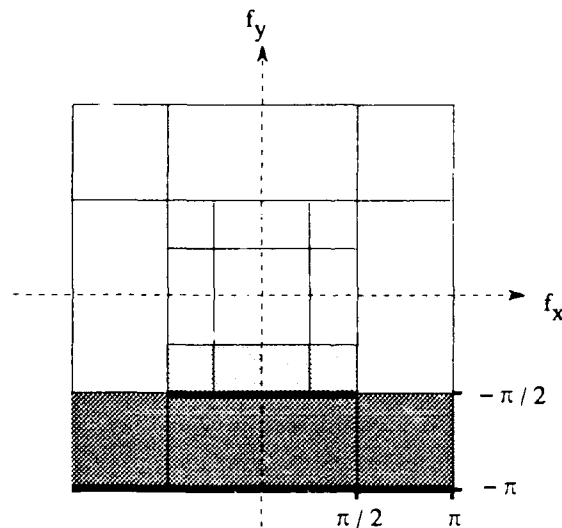


Figure 34. Frequency supports of the wavelet filters generated by several spatial decompositions for a plane taken through the FFT of the image sequence in Figure 33a) at the temporal frequency $f_t = \pi$. The dark lines represent the 2D projections of the Fourier transforms of the moving objects onto the $f_t = \pi$ frequency plane.

filters were applied at each resolution level to capture either horizontal or diagonal object features. Additionally, since Figure 35b) was obtained by decomposing the image sequence two levels in space, the dimensions of each image in Figure 35b) are now one-quarter the size of the images in the original $64 \times 64 \times 64$ image sequence. Finally, as before, a threshold was applied to capture the largest coefficients (in magnitude) at each resolution level.

The third sequence of images, shown in Figure 36, was chosen to demonstrate the motion-oriented algorithm's ability to zoom-in and zoom-out on spatial and temporal details in a natural image sequence. The image sequence contains a large, slow moving tank executing a 180° turn. The imagery is corrupted by background noise, and a plume of hot gasses is evident behind the tank in frame 100 after it executes the turn. In addition, the image jitters slightly from frame to frame, presumably as a result of slight movements in the camera platform. The image dimensions are 128×128 (row, columns, frames) and the values are eight bit grayscale (0 through 255).

The tank is fairly large and its movement is slow compared to other objects in the scene (notably, the rapidly changing pixels associated with background noise). Thus, a large amount of its energy should be contained in the coefficients corresponding to wavelets with longer dilations (i.e., lower resolutions) in space and time. This behavior is clearly evident in Figure 37 which contains a single frame from each of several different decomposition levels in space and time. Moving horizontally from right to left across the top of figure, which corresponds to decreasing the spatial resolution for a fixed temporal resolution, it is evident that the energy in the spatial wavelet coefficients increases. However, since the temporal resolution is held constant at the highest level, the scintillating background pixels with their correspondingly high temporal frequency energy are still very much present in the scene. If one now moves vertically down the right side of the figure, so that the temporal resolution decreases for a fixed spatial resolution, the coefficients associated with the large, slow tank become more and more evident, until, at the spatio-temporal resolution level $j = -3, k = -2$, only the tank remains in the image. Thus, a wavelet filter tuned to large, moderately slow moving objects successfully extracts the tank from the noisy image sequence.

A second test conducted on the tank image sequence was performed to compare the motion segmentation properties of the motion-oriented multiresolution analysis with a more traditional segmentation technique known as frame differencing. In this technique, pixel values in an image frame at time $t + 1$ are subtracted from the image at time t in order to remove stationary objects from the scene.

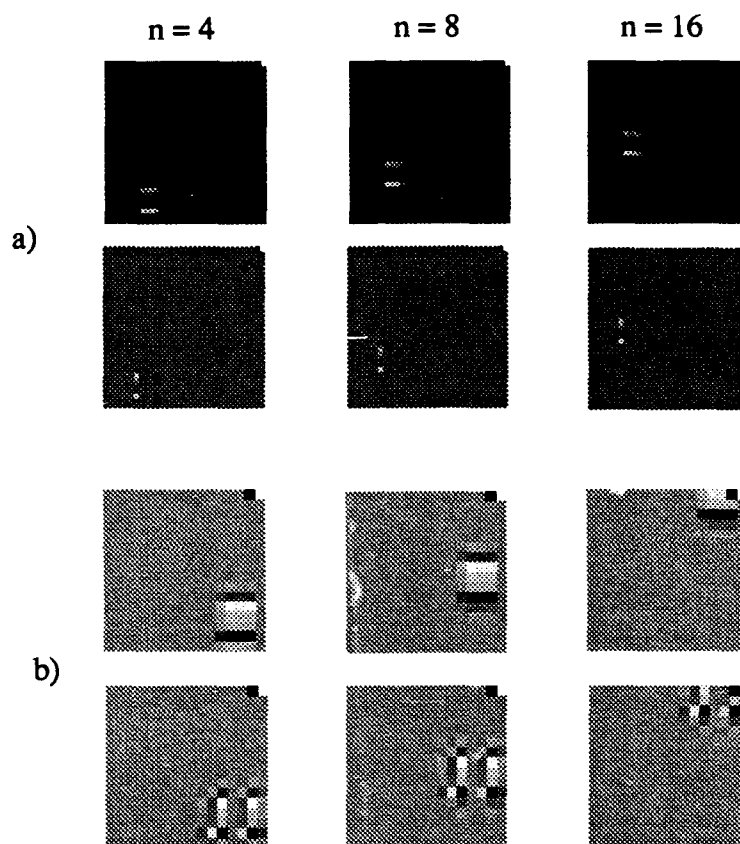


Figure 35. a) Segmenting the diagonal and horizontal features of the larger, faster object by decomposing the input signal two levels in space and one level in time. Smaller, slower object contained in original input image sequence is completely removed from the scene. The coefficient sequence dimensions are $16 \times 16 \times 32$. b) Segmenting the smaller, slower object by decomposing one level in space and one level in time. In this case, the larger, faster object has been eliminated by the motion-oriented filter bank. The coefficient sequence dimensions are $32 \times 32 \times 32$.

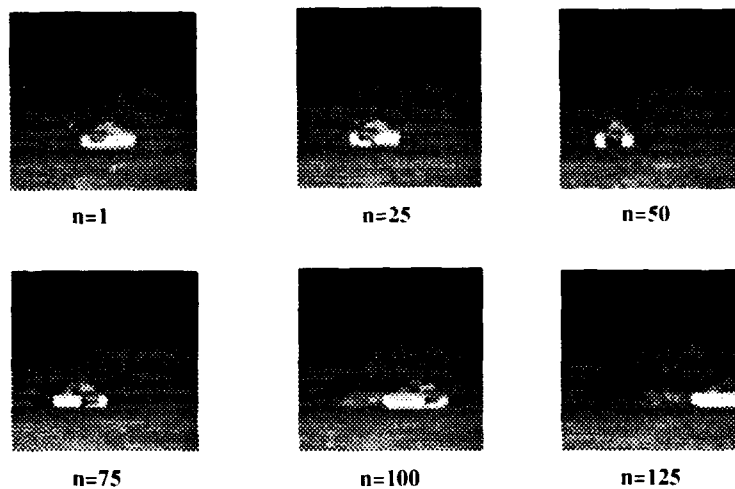


Figure 36. Several frames of a sequence of IR images in which a large, slow moving tank executes a 180° turn.

This technique is used extensively in real-time motion detection systems such as the multiresolution Pyramid Vision Machine developed by P. Burt (10). Two common problems with frame differencing techniques are 1) they require pixel registration between image frames in order to “subtract out” stationary information, and 2) frame to frame pixel scintillations caused by noise are not removed by the differencing process. The major advantage of the technique is that it can be implemented in real time. Figure 38 compares several unprocessed frames of wavelet detail coefficients to similar poses of the tank produced by a simple frame-differencing operation.

Figure 38b) shows the wavelet coefficients at the spatio-temporal resolution level $j = 1, k = 3$. In each frame, the spatial and temporal detail signals of the vertical, horizontal and diagonal features are combined to yield a complete outline of the tank. The motion-oriented wavelet decomposition algorithm has captured the edges of the tank while virtually eliminating the noisy background. The frame differencing technique, shown in Figure 38a), also captures edge information, however this technique is clearly more susceptible to noise sources in the “stationary” background. In Burt’s smart sensing pyramid vision scheme, objects of interest in the image sequence are located by analyzing frame differenced images at multiple spatial resolutions. However, Figures 37 and 38 show these objects can be significantly obscured by noise and other motion related phenomena (such as camera jitter) in the scene. The motion-oriented decomposition tool, on the other hand, allows one to independently

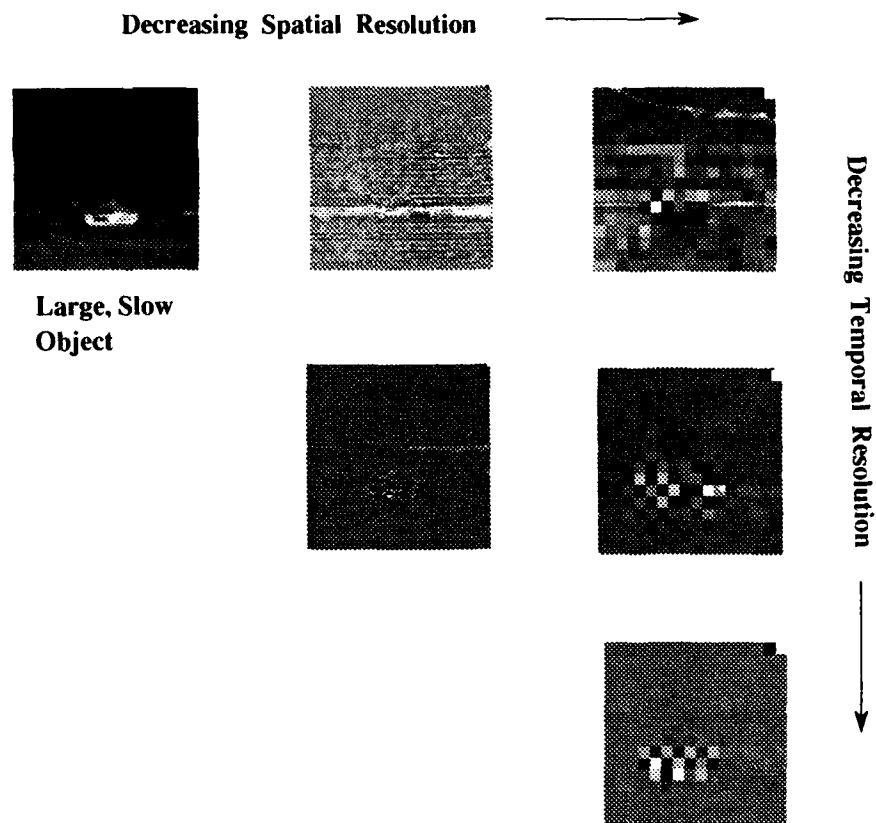


Figure 37. Single, unprocessed frame of detail coefficients from each of several different motion decompositions in space and time. Moving left to right across the figure increases the spatial dilation of the wavelet which in turn extracts lower spatial frequencies from the sequence. Moving from top to bottom increases the wavelet's temporal dilation, thereby extracting lower temporal frequencies from the sequence. The lower right image corresponds to a spatio-temporal resolution of $j = -3, k = -2$.

frame

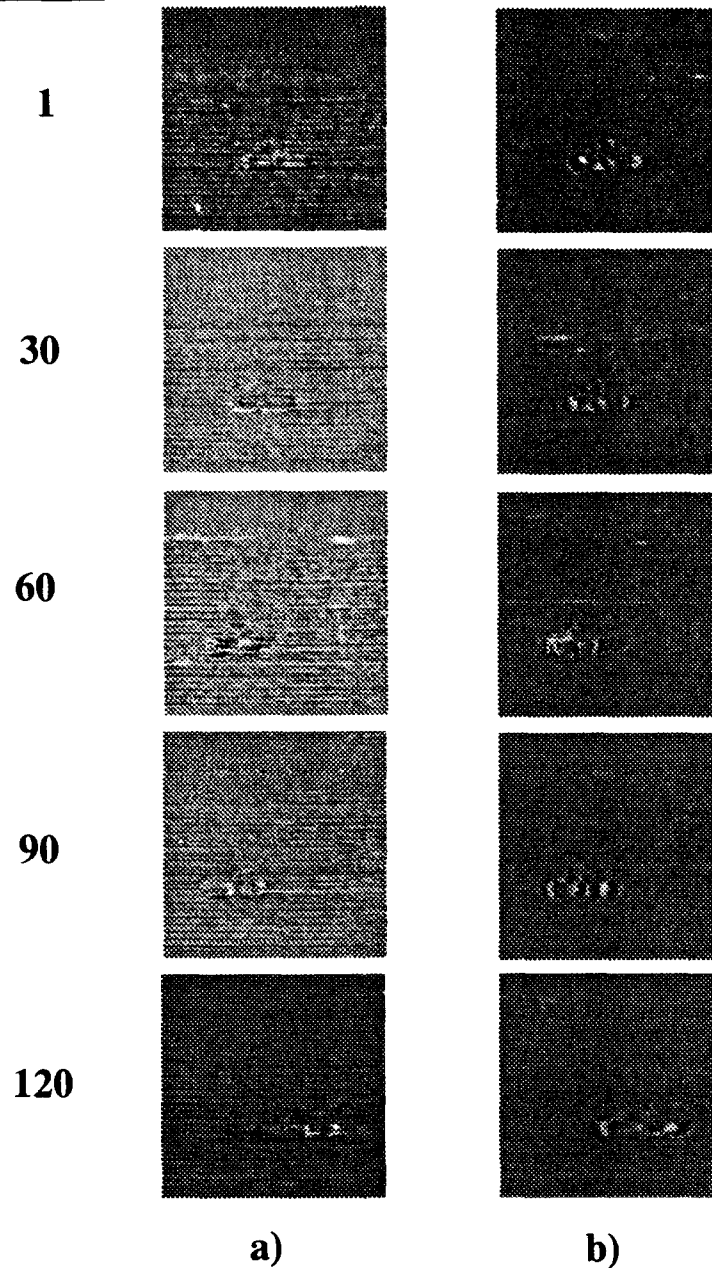


Figure 38. a) Several frames of moving tank image sequence processed with a traditional frame-differencing motion extraction technique. b) Detail coefficients generated by a motion-oriented wavelet decomposition at the spatial and temporal resolution levels $j = 1, k = 3$. The temporal details of the horizontal, vertical and corner spatial details have been combined to form an outline of the moving object.

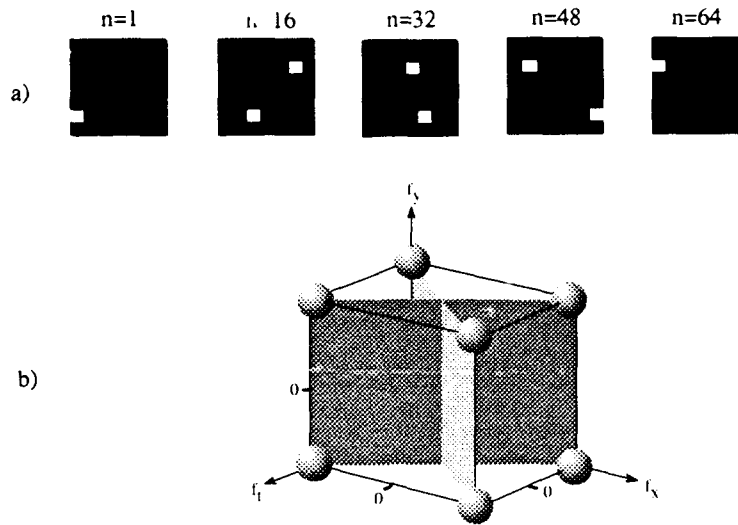


Figure 39. a) Several frames of 64×64 imagery containing two identical rectangles traveling with the velocity components $v_y = 0$ frame/sec. and $v_x = \pm 1$ frame/sec. b) A visualization of the planes containing the Fourier transforms of both rectangles. The spheres represent the frequency support of one of the seven “detail” wavelet filters.

examine spatial *and* temporal details in an image sequence in order to simultaneously locate features at different scales and eliminate extraneous motion related information.

The final test conducted in this section is designed to reveal the directional insensitivity of the motion-oriented algorithm as discussed in Section 3.6. The image sequence used for this test consists of two identical rectangles moving horizontally across the field of view at the same speeds but in opposite directions. The y velocity component of both objects equals zero and the x velocity components are $v_x = \pm 1$ frame/sec. Several frames of the moving objects, as well as the planes containing their Fourier transforms are shown in Figure 39. The dimensions of the discrete image volume are $64 \times 64 \times 64$.

The spheres in all eight corners of the frequency volume shown in Figure 39b) represent the frequency support of one of the seven detail filters generated by one step in the spatial and temporal decomposition process (i.e., $j = k = -1$). Since the coefficients of the QMF pair that generate the filter are real, the frequency supports are symmetric about all three axes. The Fourier transforms of both objects lie on the planes that cut through diagonally opposing quadrants of the frequency volume. Although the planes correspond to motion in two opposite directions, they both pass through four of the eight spheres. Clearly, this filter will “ring” in the presence of either of the two moving

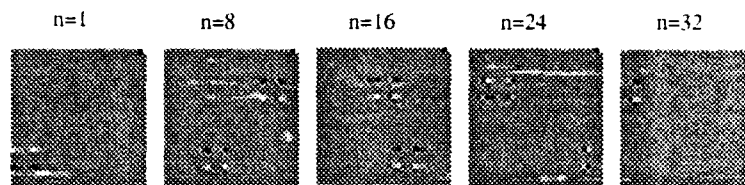


Figure 40. The “diagonal” detail coefficients obtained by one step in the motion-oriented spatial and temporal decomposition processes. The filter associated with these coefficients captures both objects even though they are traveling in opposite directions.

objects. Indeed, as Figure 40 shows, the coefficients associated with this wavelet filter capture both oppositely moving rectangles. Thus, the detail filters generated by the motion-oriented decomposition algorithm respond more like *scalar* motion (or speed) detectors than *vector* motion detectors. In order to increase the directional selectivity of the wavelet filter, Chapter V employs a Hilbert transform decomposition technique that allows one to capture the energy contained only in diagonally opposing filter pairs (e.g., the two spheres located in the opposing octants defined by $f_x > 0, f_y > 0, f_t > 0$ and $f_x < 0, f_y < 0, f_t < 0$).

4.6 Conclusions

This chapter presented an unconventional $L_2(\mathbb{R}^3)$ multiresolution wavelet analysis designed for the purpose of analyzing motion in time sequential imagery. A theoretical framework was first developed that allows for the construction of an $L_2(\mathbb{R}^3)$ multiresolution wavelet analysis from three non-identical $L_2(\mathbb{R})$ spatial and temporal multiresolution wavelet analyses. This framework provides greater flexibility for tailoring the spatio-temporal frequency characteristics of the three dimensional wavelet filter to match the frequency behavior of the analyzed signal. An unconventional, discrete multiresolution wavelet decomposition algorithm was then described which yields a rich set of independent *spatio-temporally* oriented frequency channels for analyzing the size and speed characteristics of moving objects. Unlike the conventional $L_2(\mathbb{R}^3)$ wavelet decomposition method described in Chapter III, this “motion-oriented” algorithm provides independent zoom-in and zoom-out capability in space and time.

The motion oriented algorithm was applied to a natural image sequence and several synthetic image sequences in order to demonstrate its capabilities and limitations. It was shown that decoupling

the conventional spatial and temporal decomposition processes provides the ability to segment objects with dissimilar spatial and temporal frequency characteristics (e.g., small, slow objects or large, fast objects). Additionally, the independent zoom-in and zoom-out capability of the motion-oriented decomposition tool allows one to locate objects at different spatial scales in the presence of extraneous motion related phenomena such as camera jitter, background noise and sensor noise.

The final example demonstrated that the motion-oriented algorithm produces scalar motion sensors that are sensitive to object speed, but insensitive to object direction. In Chapter V, a Hilbert Transform is used in conjunction with the unconventional wavelet decomposition process to produce vector motion sensors that respond preferentially to vertical, horizontal or diagonal features corresponding to a given object's size, speed, direction and location in the scene.

V. Object Discrimination Using a Motion-Oriented Wavelet Multiresolution Analysis

5.1 Introduction

The previous chapter presented a unique motion-oriented $L_2(\mathbb{R}^3)$ multiresolution wavelet analysis designed to detect objects of different sizes moving with different speeds across a two-dimensional image plane. Furthermore, it was shown that the symmetric wavelet detail filters generated by the motion-oriented wavelet analysis act as scalar motion sensors in that they respond to the magnitude of an object's velocity vector (i.e., its speed), rather than to the vector quantity of speed and direction. The purpose of this chapter, therefore, is to expand the properties of the motion-oriented wavelet analysis to provide a multiresolution motion analysis tool that discriminates multiple moving objects in a three-dimensional image sequence based on their location, size, speed and direction of motion. The chapter is divided into two major sections. The first section provides the mathematical foundation for the vector motion analysis tool by combining the properties of the Hilbert transform with the motion-oriented multiresolution wavelet analysis to yield a bank of directionally selective wavelet filters. An algorithm is then presented which combines the responses of the directionally selective wavelet filters to discriminate multiple objects in a 3D image sequence by computing the optical flow. The second major section of the chapter introduces a unique cooperative-competitive strategy that restores localized flow fields corrupted by noise. The strategy employs a modified gated dipole filter designed to reinforce consistent flow behavior and remove flow inconsistencies. Several examples are provided which demonstrate the utility of the gated dipole flow restoration process.

5.2 A Vector Wavelet Motion Sensor

Section 4.5 presented several examples which demonstrated the capabilities of the discrete motion-oriented multiresolution wavelet analysis. These included the ability to differentiate between objects moving in an image sequence based on their location, size and speed. The final example, however, served to emphasize a key limitation of the motion analysis technique - it is not selective for motion direction. Since velocity is a vector quantity consisting of both speed and direction, this limitation constitutes a serious shortcoming for a motion analysis tool.

The inability of the motion-oriented multiresolution wavelet analysis to respond preferentially to direction of motion is attributable to the symmetry of the quadrature mirror filters in three-dimensional

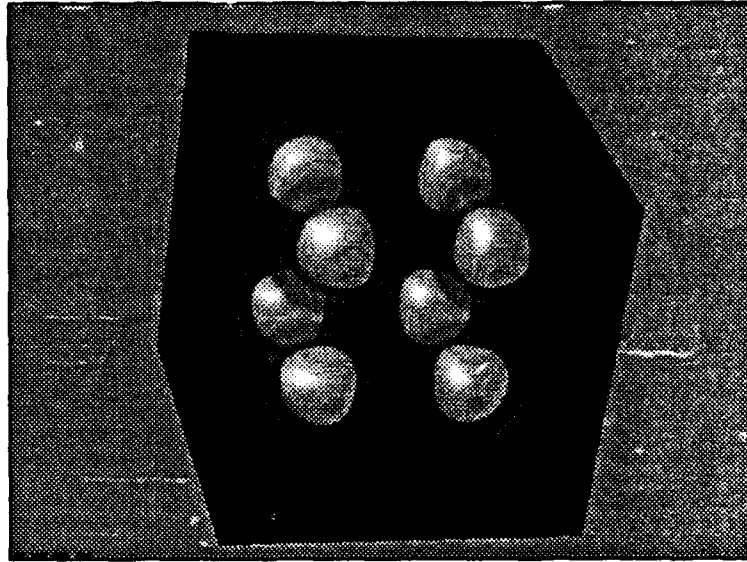


Figure 41. 3D Fourier transform of the wavelet $\Psi_j^7(x, y, t)$ graphically rendered using a ray tracing program developed at AFIT.

frequency space. This symmetry is clearly evident in Figure 41 which shows a three-dimensional Fourier transform of the “seventh” detail wavelet, $\Psi_j^7(x, y, t)$ as defined in Theorem 1. The wavelet was iteratively constructed from Daubechies 7 g filters in space and time. The figure was obtained by graphically rendering the Fourier transform of $\Psi_j^7(x, y, t)$ using a ray tracing program developed at AFIT (34).

As discussed in Section 4.5, when two objects move horizontally in opposite directions and at equal speeds across an image plane, their Fourier transforms will lie on two planes as shown in Figure 42a). Assuming the speed of the objects matches the temporal frequency characteristics of the Fourier transform in Figure 41, it's clear the corresponding filter will capture both objects in the image sequence. One might think that the inability to preferentially discriminate one object from the other might be overcome by devising a way to cancel the response of the filter in four of the regions contained in opposite quadrants of the frequency volume as shown by the four black spheres in Figure 42a). One could then selectively extract objects moving in either direction by canceling the response of the filter regions in the appropriate quadrants. The problem with this approach, however, is demonstrated by the frequency response shown in Figure 42b) which contains the planes associated with two objects, one moving vertically *and* the other moving horizontally at the same speeds.

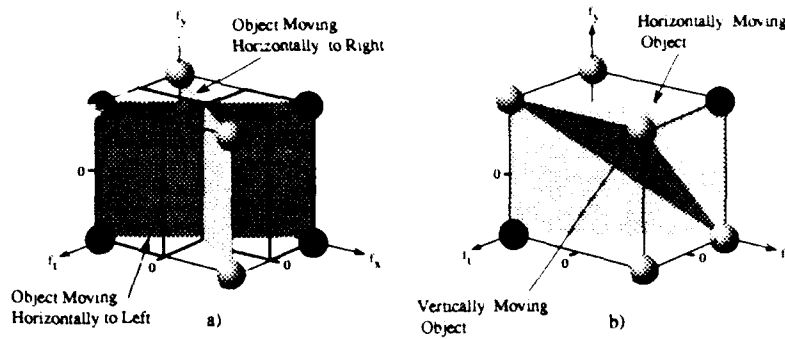


Figure 42. a) Fourier transforms of two objects moving horizontally with the same speed but in opposite directions. Opposing *quadrants* are highlighted in dark lines. The spheres depict the frequency supports of the filter shown in Figure 41. b) Fourier transforms of two objects traveling horizontally and vertically with equal speeds. The black spheres are located in diagonally opposing *octants* of the frequency volume.

Figure 42b) shows that by selectively canceling the filter response only in opposing *quadrants* of the frequency volume, one cannot preferentially segment either of the horizontally and vertically traveling objects. However, if it were possible to cancel the response of the filter everywhere but in diagonally opposing *octants* of the frequency volume, as shown by the black spheres in Figure 42b), one could theoretically discriminate between both objects in the image sequence (58). This section describes a method for obtaining such a frequency response through the use of the Hilbert transform. Although the Hilbert transform was employed for a similar purpose by Watson and Ahumada (58) the contribution in this phase of the research consists of the creation of an "extended" real signal which is incorporated in the motion-oriented multiresolution wavelet analysis to yield diagonally opposing wavelet filters at all possible spatial and temporal resolutions. Several properties of the Hilbert transform that are relevant to this objective are discussed next.

5.2.1 The Hilbert Transform. The Hilbert transform is a convolution operator with the transfer function, $Hil(f)$, where (60)

$$Hil(f) = -j\text{sgn}(f) \quad (116)$$

The sgn or “signum” function is defined as

$$\text{sgn}(f) = \begin{cases} 1, & f > 0 \\ 0, & f = 0 \\ -1, & f < 0 \end{cases} \quad (117)$$

so that the effect of the Hilbert transform is to shift the phase of all frequency components of a signal by a factor of $-\frac{\pi}{2}$ radians. If $X(f)$ is the Fourier transform of an input signal $x(t)$, then its Hilbert transform, $\hat{x}(t)$, is

$$\begin{aligned} \hat{x}(t) &= \mathcal{F}^{-1}\{X(f)\text{Hil}(f)\} \\ &= x(t) * \text{hil}(t) \end{aligned} \quad (118)$$

where the impulse response, $\text{hil}(t) = -j\mathcal{F}^{-1}\{\text{sgn}(f)\}$, is obtained from the Fourier transform pair

$$\frac{j}{\pi t} \leftrightarrow \text{sgn}(f) \quad (119)$$

and where the double headed arrow denotes the Fourier and inverse Fourier transform operations. Inserting Equation 119 into Equation 118 then yields the following definition for the Hilbert transform of $x(t)$

$$\begin{aligned} \hat{x}(t) &= \int_{-\infty}^{\infty} \frac{x(\tau)}{\pi(t - \tau)} d\tau \\ &= \int_{-\infty}^{\infty} \frac{x(t - \tau)}{\pi\tau} d\tau \end{aligned} \quad (120)$$

Clearly, $\hat{x}(t)$ cannot be computed at $t = 0$, however, in this research, the Hilbert transform is implemented in the Fourier domain where this problem does not arise (58).

Now consider the complex analytic signal, $x_a(t)$, defined in terms of the real signal $x(t)$ and its Hilbert transform, $\hat{x}(t)$,

$$x_a(t) = x(t) + j\hat{x}(t) \quad (121)$$

The spectrum of the analytic signal is often used in single-sideband communication systems to reduce the transmission bandwidth of a signal (60). It is obtained by computing the Fourier transform of the

right hand side of Equation 121 as follows

$$\begin{aligned}
 X_a(f) &= X(f) + j\hat{X}(f) \\
 &= X(f) + jH(f)X'(f) \\
 &= X(f) + j[-j\text{sgn}(f)X(f)] \\
 &= X(f)[1 + \text{sgn}(f)]
 \end{aligned} \tag{122}$$

From the definition of the signum function, Equation 122 can be expressed as

$$X_a(f) = \begin{cases} 2X(f), & f > 0 \\ 0, & f < 0 \end{cases} \tag{123}$$

which shows the frequency response of the analytic signal contains only the positive frequency components in the signal's spectrum. Similarly, by changing the sign in the first line of Equation 122 from a positive to a negative sign, one can retain only the negative frequency components of the signal spectrum. This property is employed in the following section to selectively choose diagonally opposing octants in 3D frequency space.

5.2.2 Directionally Selective Wavelet Filters. As a first step in understanding how the Hilbert transform can be used to generate the directionally selective filters described above, consider again the dyadic Wavelet transform of the one-dimensional signal $x(t)$ previously presented in Chapter II,

$$[Wf](2^l, m) = 2^{-\frac{l}{2}} \int_{-\infty}^{\infty} x(t) \psi(2^l t - m) dt \tag{124}$$

where l and m are integers that specify the dilation and translation of the wavelet kernel and $\psi(t)$ is a mother wavelet. As in Equation 4 in Section 2.2.1, the wavelet transform can be expressed in terms of the convolution integral

$$\begin{aligned}
 [Wf](2^l, m) &= \int_{-\infty}^{\infty} x(t) \psi_l(m - t) dt \\
 &= x(t) * \psi_l(t)
 \end{aligned} \tag{125}$$

where $\psi_l(t) = 2^{-\frac{l}{2}} \psi(-2^l t)$. Next, define the pair of *analytic mother wavelets* by

$$\psi_a^\pm(t) = \psi(t) \pm j\hat{\psi}(t) \quad (126)$$

Notice that the analytic wavelet is indeed a “wavelet” in that it meets the admissibility condition

$$\int_{-\infty}^{\infty} \frac{|\hat{\Psi}_a^\pm(\omega)|^2}{|\omega|} d\omega < \infty \quad (127)$$

given in Section 2.2.1. This follows from the fact that the energy in a signal and its Hilbert transform are equal. That is,

$$|\hat{\Psi}(\omega)|^2 \equiv |\mathcal{F}\{\hat{\psi}(t)\}|^2 = |-j \operatorname{sgn} \omega|^2 |\Psi(\omega)|^2 = |\Psi(\omega)|^2 \quad (128)$$

Using the relationship derived in Equation 122, the spectrum of the analytic mother wavelet pair can be expressed by

$$\Psi_a^+(f) = \begin{cases} 2\Psi(f), & f > 0 \\ 0, & f < 0 \end{cases} \quad (129)$$

and

$$\Psi_a^-(f) = \begin{cases} 0, & f > 0 \\ 2\psi(f), & f < 0 \end{cases} \quad (130)$$

where the symbols “+” and “−” indicate the positive and negative halves of the frequency axis and $\Psi(f)$ is the Fourier transform of $\psi(t)$. The moduli of both “one-sided” spectra for a Daubechies 12 analytic wavelet pair are shown in Figure 43.

Now define the analytic wavelet *transform* pair by

$$\begin{aligned} [W_a^\pm x](2^l, m) &= 2^{-\frac{l}{2}} \int_{-\infty}^{\infty} x(t) \psi_a^\pm(2^l t - m) dt \\ &= 2^{-\frac{l}{2}} \left(\int_{-\infty}^{\infty} x(t) \psi(2^l t - m) dt \pm j \int_{-\infty}^{\infty} x(t) \hat{\psi}(2^l t - m) dt \right) \end{aligned} \quad (131)$$

where $\hat{\psi}(t)$ is the Hilbert transform of the wavelet $\psi(t)$. Using the previously defined substitution variable $\psi_l(t)$ and the new variable $\hat{\psi}_l(t) = 2^{-\frac{l}{2}} \hat{\psi}(-2^l t)$, Equation 131 can be expressed as the sum

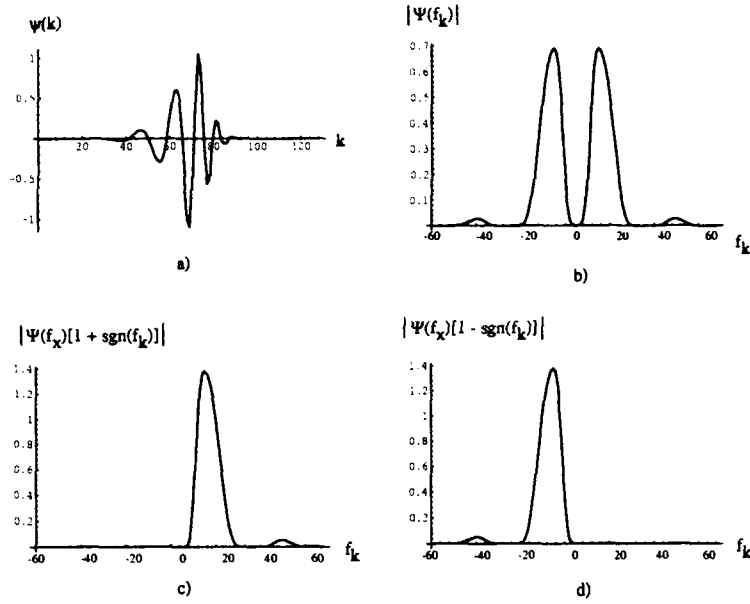


Figure 43. a) A Daubechies 12 wavelet and b) the magnitude of its Fourier transform. c) The one-sided spectrum of the analytic signal $\psi_a^+(t) = \psi(t) + j\hat{\psi}(t)$ and d) the one-sided spectrum of the analytic signal $\psi_a^-(t) = \psi(t) - j\hat{\psi}(t)$.

of two convolution integrals:

$$\begin{aligned}
 [W_a^\pm x](2^l, m) &= \int_{-\infty}^{\infty} x(t) \psi_l(m-t) dt \pm j \int_{-\infty}^{\infty} x(t) \hat{\psi}_l(m-t) dt \\
 &= x(t) * \psi_l(t) \pm j x(t) * \hat{\psi}_l(t) \\
 &= x(t) * \psi_l(t) \pm j x(t) * [hil(t) * \psi_l(t)] \\
 &= x(t) * \psi_l(t) \pm j [x(t) * hil(t)] * \psi_l(t) \\
 &= x(t) * \psi_l(t) \pm j \hat{x}(t) * \psi_l(t) \\
 &= x_a(t) * \psi_l(t)
 \end{aligned} \tag{132}$$

Equation 132 shows that the analytic wavelet transform of a real signal, $x(t)$ is equivalent to the wavelet transform of the analytic signal, $x_a(t)$ where $x_a(t) = x(t) + j\hat{x}(t)$. Since the transforms are obtained using convolution integrals, this process can also be implemented in the Fourier frequency domain as follows,

$$[W_a^\pm x](2^l, m) = x(t) * \psi_l(t) \pm j \hat{x}(t) * \psi_l(t)$$

$$\begin{aligned}
&= \mathcal{F}^{-1}\{X(f)\Psi_i(f) \pm j[-j\text{sgn}(f)X(f)\Psi_i(f)]\} \\
&= \mathcal{F}^{-1}\{X(f)\Psi_i(f)[1 \pm \text{sgn}(f)]\}
\end{aligned} \tag{133}$$

Viewed as filtering operation, the “one-sided” analytic wavelet transforms capture either the positive or negative frequency components in the signal spectrum that lie within the bandpass regions of the wavelet filter $\Psi(f)$. The transforms are easily obtained by filtering the input signal with $\Psi(f)$ and retaining only the positive or negative portions of the frequency spectrum. Since the resulting spectrums are one-sided (i.e., asymmetric), the one-dimensional analytic wavelet transforms will always be complex. Next consider how one can use a Hilbert transform in conjunction with a three dimensional Wavelet transform to capture the frequency components contained in diagonally opposing octants of a three-dimensional frequency spectrum.

In order to obtain directional selectivity in a spatio-temporal frequency analysis, the previous section showed a filter is required that possesses identical regions of support in diagonally opposing octants in three-dimensional frequency space. Because of its symmetry, the 3D wavelet filter yields identical regions of support in all eight frequency quadrants. It is possible, however, to capture any two diagonally opposing wavelet filter regions through the judicious application of multiple 1D Hilbert transforms. For example, consider the “extended *real* mother wavelet” given by

$$\begin{aligned}
\psi_{r18}(x, y, t) = & \psi(x, y, t) - \text{hil}(x) * \text{hil}(y) * \psi(x, y, t) - \text{hil}(x) * \text{hil}(t) * \psi(x, y, t) \\
& - \text{hil}(y) * \text{hil}(t) * \psi(x, y, t)
\end{aligned} \tag{134}$$

where the subscript *r18* indicates the extended mother wavelet is *real* and, as shown next, its Fourier transform captures frequencies in the first and eighth diagonally opposing octants of 3D frequency space as defined in Figure 44.

Once again using the Hilbert transform pair

$$\frac{j}{\pi t} \leftrightarrow \text{sgn}(f) \tag{135}$$

the Fourier transform of $\psi_{r18}(x, y, t)$ can be written as

$$\mathcal{F}\{\psi_{r18}(x, y, t)\} = \Psi(f_x, f_y, f_t)[1 + \text{sgn}(f_x)\text{sgn}(f_y) + \text{sgn}(f_x)\text{sgn}(f_t) + \text{sgn}(f_y)\text{sgn}(f_t)]$$

octant	f_x	f_y	f_t	Ψ_{r18}
1	+	+	+	4Ψ
2	+	-	+	0
3	-	+	+	0
4	-	-	+	0
5	+	-	-	0
6	+	+	-	0
7	-	+	-	0
8	-	-	-	4Ψ

Figure 44. Table showing the frequency response of the extended wavelet $\psi_{r18}(x, y, t)$ in each of the eight octants of a spatio-temporal frequency volume. The plus and minus signs indicate the corresponding frequency region is greater than or less than zero respectively. The frequency response is non-zero in the diagonally opposing octants 1 and 8 where it is four times the frequency response of the constructing wavelet $\psi(x, y, t)$.

$$= \Psi(f_x, f_y, f_t)[1 + \text{sgn}(f_x, f_y) + \text{sgn}(f_x, f_t) + \text{sgn}(f_y, f_t)] \quad (136)$$

where the real, separable 2D signum function, $\text{sgn}(f_u, f_v)$, is formed from the product of the imaginary 1D signum functions $-j\text{sgn}(f_u) \cdot -j\text{sgn}(f_v)$. The table in Figure 44 shows the frequency response of Equation 136 for each of the eight octants in a spatio-temporal frequency volume. Evidently the frequency response is non-zero only in the first and the eighth octants, which correspond to the diagonally opposing frequency regions $f_x > 0, f_y > 0, f_t > 0$ and $f_x < 0, f_y < 0, f_t < 0$. Furthermore, the frequency response in these two regions is four times the response of the constructing wavelet $\psi(x, y, t)$. Figure 45 shows the 3D Fourier transform of the extended wavelet $\psi_{r18}(x, y, t)$, where the constructing wavelet employs a Daubechies 7 g filter in space and time. The Fourier transform of the constructing wavelet was previously shown in Figure 41. As before, the Fourier transform of $\psi_{r18}(x, y, t)$ was graphically rendered using a 3D ray-tracing program.

The frequency response of the extended wavelet $\psi_{r18}(x, y, t)$ captures the spatio-temporal frequency components lying in the diagonally opposing frequency regions $f_x > 0, f_y > 0, f_t > 0$ and $f_x < 0, f_y < 0, f_t < 0$. Three additional extended wavelets are needed to capture the remaining three diagonally opposing octants. The three wavelets and their Fourier frequency responses are shown

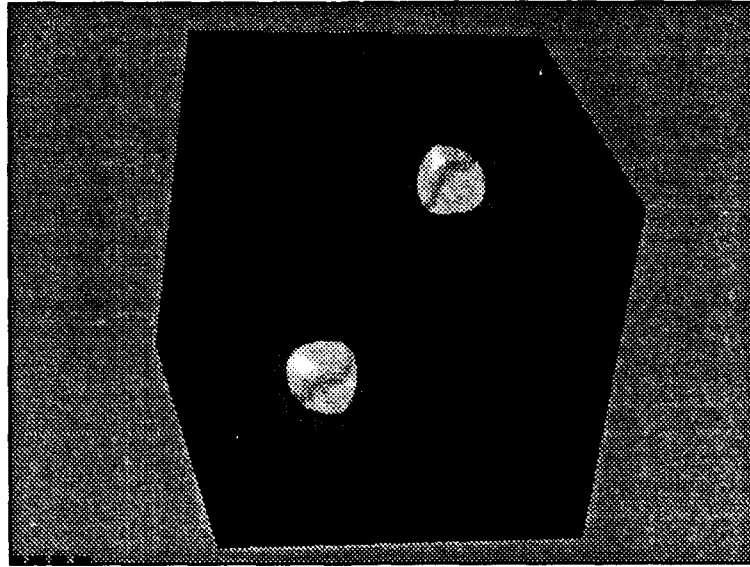


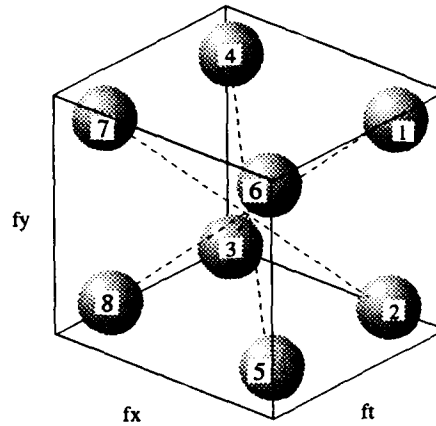
Figure 45. 3D Fourier transform of the extended wavelet $\psi_{r18}(x, y, t)$ graphically rendered using a ray tracing program developed at AFIT. The constructing wavelet employs a Daubechies 7 g filter in space and time.

in Equations 137 through 139. A visualization of the regions of support in spatio-temporal frequency space of each of the four extended wavelet's Fourier transforms is provided in Figure 46.

$$\begin{aligned}
 \psi_{r27}(x, y, t) &= \psi(x, y, t) + hil(x) * hil(y) * \psi(x, y, t) - hil(x) * hil(t) * \psi(x, y, t) \\
 &\quad + hil(y) * hil(t) * \psi(x, y, t) \\
 \mathcal{F}\{\psi_{r27}(x, y, t)\} &= \Psi(f_x, f_y, f_t)[1 - \text{sgn}(f_x, f_y) + \text{sgn}(f_x, f_t) - \text{sgn}(f_y, f_t)] \quad (137)
 \end{aligned}$$

$$\begin{aligned}
 \psi_{r36}(x, y, t) &= \psi(x, y, t) - hil(x) * hil(y) * \psi(x, y, t) + hil(x) * hil(t) * \psi(x, y, t) \\
 &\quad + hil(y) * hil(t) * \psi(x, y, t) \\
 \mathcal{F}\{\psi_{r36}(x, y, t)\} &= \Psi(f_x, f_y, f_t)[1 + \text{sgn}(f_x, f_y) - \text{sgn}(f_x, f_t) - \text{sgn}(f_y, f_t)] \quad (138)
 \end{aligned}$$

$$\begin{aligned}
\psi_{r45}(x, y, t) &= \psi(x, y, t) + \text{hil}(x) * \text{hil}(y) * \psi(x, y, t) + \text{hil}(x) * \text{hil}(t) * \psi(x, y, t) \\
&\quad - \text{hil}(y) * \text{hil}(t) * \psi(x, y, t) \\
\mathcal{F}\{\psi_{r45}(x, y, t)\} &= \Psi(f_x, f_y, f_t)[1 - \text{sgn}(f_x, f_y) + \text{sgn}(f_x, f_t) + \text{sgn}(f_y, f_t)] \quad (139)
\end{aligned}$$



- 18: $\Psi(f_x, f_y, f_t)(1 + \text{sgn}(f_x)\text{sgn}(f_y) + \text{sgn}(f_x)\text{sgn}(f_t) + \text{sgn}(f_y)\text{sgn}(f_t))$
- 27: $\Psi(f_x, f_y, f_t)(1 - \text{sgn}(f_x)\text{sgn}(f_y) + \text{sgn}(f_x)\text{sgn}(f_t) - \text{sgn}(f_y)\text{sgn}(f_t))$
- 36: $\Psi(f_x, f_y, f_t)(1 + \text{sgn}(f_x)\text{sgn}(f_y) - \text{sgn}(f_x)\text{sgn}(f_t) - \text{sgn}(f_y)\text{sgn}(f_t))$
- 45: $\Psi(f_x, f_y, f_t)(1 - \text{sgn}(f_x)\text{sgn}(f_y) - \text{sgn}(f_x)\text{sgn}(f_t) + \text{sgn}(f_y)\text{sgn}(f_t))$

Figure 46. A visualization of the regions of support in spatio-temporal frequency space of the Fourier transforms of the four extended wavelets $\psi_{r18}(x, y, t)$ through $\psi_{r45}(x, y, t)$ where the constructing wavelet $\psi(x, y, t)$ is bandpass in f_x, f_y and f_t . Note that each wavelet captures two diagonally opposing regions in frequency space and that four wavelets are needed to cover all eight octants.

Taken together, the four extended wavelets cover all four possible diagonally opposing regions in spatio-temporal frequency space. As argued earlier, this adds a degree of directional selectivity not provided by the symmetric frequency spectrum associated with a conventional 3D wavelet (although, it will be shown later in this chapter that full directional selectivity is obtained only by combining the responses of two extended wavelets). In the above examples, the constructing wavelet possessed the highpass spatio-temporal frequency characteristics associated with the detail wavelet $\Psi_j^7(x, y, t)$ (see Theorem 4). The following section incorporates the “extended wavelet” concept into the motion-oriented multiresolution wavelet analysis developed in Chapter IV to yield a bank of diagonally opposing wavelet filters tuned to multiple object sizes, speeds and directions.

5.2.3 Directionally Selective, Motion-Oriented Multiresolution Wavelet Analysis. The preceding section showed that one can capture diagonally opposing supporting regions of a 3D symmetric wavelet filter through the use of an extended real wavelet filter. It was also shown that the symmetric filter bank generated by the discrete motion-oriented multiresolution wavelet analysis serves as a scalar motion sensor in that it can sense the speed of moving objects but not their direction. The purpose of this section is to wed the two concepts to form the foundation for a vector motion sensing tool that responds preferentially to a given object size, speed and direction.

Begin the development by assuming that $f(x, y, t)$ represents an $L_2(\mathbb{R}^3)$ spatio-temporal signal. Without loss of generality, construct the extended real signal, $f^e(x, y, t)$, as follows

$$\begin{aligned} f^e(x, y, t) = & f(x, y, t) - hil(x) * hil(y) * f(x, y, t) - hil(x) * hil(t) * f(x, y, t) \\ & - hil(y) * hil(t) * f(x, y, t) \end{aligned} \quad (140)$$

where $hil(x)$ is the Hilbert transform integral kernel given in Section 5.1.1. and the signs in Equation 140 match those given in Equation 134. Theorem 8 shows that $f^e(x, y, t)$ is an element of $L_2(\mathbb{R}^3)$ and can therefore be decomposed under a wavelet multiresolution analysis.

Theorem 8. Let $f \in L_2(\mathbb{R}^3)$. Then f^e is also contained in $L_2(\mathbb{R}^3)$.

Proof. Let $g(x, y, t) = \text{hil}(x) * \text{hil}(y) * f(x, y, t)$, and show that $g \in L_2(\mathbb{R}^3)$. That is, show $\|g(x, y, t)\|^2 < \infty$ where $\|\cdot\|$ indicates the L_2 norm. Now, by Parseval's Identity,

$$\begin{aligned}\|g(x, y, t)\|^2 &= \|\mathcal{F}\{g(x, y, t)\}\|^2 \\ &= \|(-j\text{sgn}(\omega_x))(-j\text{sgn}(\omega_y))F(\omega_x, \omega_y, \omega_t)\|^2 \\ &= \|F(\omega_x, \omega_y, \omega_t)\|^2\end{aligned}$$

where $F(\omega_x, \omega_y, \omega_t) = \mathcal{F}\{f(x, y, t)\}$. But, the Fourier transform maps $L_2(\mathbb{R}^3)$ onto $L_2(\mathbb{R}^3)$. Thus, $\|F(\omega_x, \omega_y, \omega_t)\|^2 < \infty$ which implies $\|g(x, y, t)\|^2 < \infty$. Furthermore, this result implies the L_2 norm of each of the last two components in $f^e(x, y, t)$ are also finite, so that

$$\begin{aligned}\|f^e(x, y, t)\|^2 &= \|f(x, y, t) - \text{hil}(x) * \text{hil}(y) * f(x, y, t) - \text{hil}(x) * \text{hil}(t) * f(x, y, t) \\ &\quad - \text{hil}(y) * \text{hil}(t) * f(x, y, t)\|^2 \\ &\leq \|f(x, y, t)\|^2 + \|\text{hil}(x) * \text{hil}(y) * f(x, y, t)\|^2 + \|\text{hil}(x) * \text{hil}(t) * f(x, y, t)\|^2 \\ &\quad + \|\text{hil}(y) * \text{hil}(t) * f(x, y, t)\|^2 \\ &< \infty\end{aligned}\tag{141}$$

Thus f^e is an element of $L_2(\mathbb{R}^3)$. **Q.E.D.**

Next, following the notation in Chapter IV, assume the projection of the extended signal onto the zeroth approximation space is given by

$$a_{0;p,q,r}^e = a_{0;p,q,r} - \text{hil}(p) * \text{hil}(q) * a_{0;p,q,r} - \text{hil}(p) * \text{hil}(r) * a_{0;p,q,r} - \text{hil}(q) * \text{hil}(r) * a_{0;p,q,r} \tag{142}$$

where p, q and r are elements in a rectangular sampling grid and $a_{0;p,q,r}$ is simply a sampled version of the original signal. Chapter IV showed that the projection coefficients at the -1 st approximation level can be obtained by discretely convolving the projection coefficients at the zeroth approximation level with the separable scaling function $\bar{h}(p)\bar{h}(q)\bar{h}(r)$ and keeping every other sample. Recall that the impulse responses $h(n)$ and $\tilde{h}(n)$ are formed from two different scaling functions. In Chapter IV,

this process was expressed by the equation

$$a_{-1;(l,m,n)} = \left[a_{0;(p,q,r)} * \bar{h}(p) * \bar{h}(q) * \bar{\bar{h}}(r) \right] (2l, 2m, 2n) \quad (143)$$

Inserting Equation 142 into Equation 143 and rearranging yields the following expression for the projection of the *extended* signal onto the $j = -1$ st approximation level

$$\begin{aligned} a_{-1;(l,m,n)}^e &= \left[a_{0;(p,q,r)}^e * \bar{h}(p) * \bar{h}(q) * \bar{\bar{h}}(r) \right] (2l, 2m, 2n) \\ &= \left[(a_{0;(p,q,r)} - hil(p) * hil(q) * a_{0;(p,q,r)} - hil(p) * hil(r) * a_{0;(p,q,r)} \right. \\ &\quad \left. - hil(q) * hil(r) * a_{0;(p,q,r)}) * \bar{h}(p) * \bar{h}(q) * \bar{\bar{h}}(r) \right] (2l, 2m, 2n) \\ &= \left[a_{0;(p,q,r)} * \{ \bar{h}(p) * \bar{h}(q) * \bar{\bar{h}}(r) - (hil(p) * hil(q)) * \bar{h}(p) * \bar{h}(q) * \bar{\bar{h}}(r) \right. \\ &\quad \left. - (hil(p) * hil(r)) * \bar{h}(p) * \bar{h}(q) * \bar{\bar{h}}(r) \right. \\ &\quad \left. - (hil(q) * hil(r)) * \bar{h}(p) * \bar{h}(q) * \bar{\bar{h}}(r) \} \right] (2l, 2m, 2n) \\ &= \left[\text{DFT}^{-1} \{ A_{0;(f_p, f_q, f_r)} \cdot \bar{H}(f_p, f_q, f_r) \{ 1 \right. \\ &\quad \left. + \text{sgn}(f_p, f_q) + \text{sgn}(f_p, f_r) + \text{sgn}(f_q, f_r) \} \} \right] (2l, 2m, 2n) \end{aligned} \quad (144)$$

where

$$\begin{aligned} \bar{H}(f_n) &= \text{DFT}\{\bar{h}(n)\} \\ \bar{\bar{H}}(f_n) &= \text{DFT}\{\bar{\bar{h}}(n)\} \end{aligned} \quad (145)$$

and

$$\bar{\bar{H}}(f_p, f_q, f_r) = \bar{H}(f_p) \bar{H}(f_q) \bar{\bar{H}}(f_r) \quad (146)$$

Equation 144 shows that the projection coefficients at the $j = -1$ approximation level for the extended signal defined in Equation 140 can be obtained by discretely filtering the originally sampled signal $a_{0;(p,q,r)}$ with the extended scaling function filter

$$\bar{\bar{H}}(f_p, f_q, f_r) \{ 1 + \text{sgn}(f_p, f_q) + \text{sgn}(f_p, f_r) + \text{sgn}(f_q, f_r) \} \quad (147)$$

inverse Fourier transforming the result, and keeping every other sample in each dimension. As discussed in the previous section, the signs in the definition of the extended signal determine the non-zero, diagonally opposing frequency regions of the extended scaling function filter. In this case, the resulting coefficients therefore represent the information in the sampled signal that lie in the passband of the scaling function filter $\bar{H}(f_p, f_q, f_r)$ contained in the diagonally opposing regions $f_p > 0, f_q > 0, f_r > 0$ and $f_p < 0, f_q < 0, f_r < 0$. The information in the remaining three diagonally opposing frequency pairs is obtained by varying the signs of the extended signal in accordance with Equations 134 through 139.

In practice, the coefficients associated with diagonally opposing frequency regions at an arbitrary spatial and temporal resolution level are computed as follows. First, components 2, 3 and 4

$$a_{0;(p,q,r)}^e = \underbrace{a_{0;(p,q,r)}}_1 - \underbrace{hil(p) * hil(q) * a_{0;(p,q,r)}}_2 - \underbrace{hil(p) * hil(r) * a_{0;(p,q,r)}}_3 - \underbrace{hil(q) * hil(r) * a_{0;(p,q,r)}}_4 \quad (148)$$

that comprise the discrete, extended real signal in Equation 148 are formed by multiplying the FFT of the sampled image sequence by the appropriate combination of 1D Hilbert transform transfer functions and inverse Fourier transforming the result. Causality problems associated with the temporal Hilbert transform are avoided by defining the mid-frame in the $N \times N \times N$ image sequence (i.e., $f(p, q, \frac{N}{2})$) as $t = 0$. After constructing the extended real signal, all four components are then decomposed individually using the discrete motion-oriented multiresolution wavelet analysis in Chapter IV. At each level of decomposition in space and time, the four sets of projection coefficients (one set for each of the extended signal components) are summed in accordance with the sign conventions in Equations 134 through 139. Each sign convention captures the information in one of four diagonally opposing frequency regions for a symmetric wavelet filter generated at a given spatial and temporal decomposition level. This process is illustrated by the flow diagram shown in Figure 47 in which the extended j th approximation coefficients are decomposed into four sets of d^j detail coefficients which are then summed to extract the information contained in diagonally opposing wavelet filters.

The “directionally selective motion-oriented multiresolution wavelet analysis” described above produces a bank of spatio-temporal filters that are selective for both object speed and, to some degree, direction of motion. The degree to which the extended wavelet filters are directionally selective is determined by the orientation of the plane in Fourier space that contains the spatio-temporal frequency

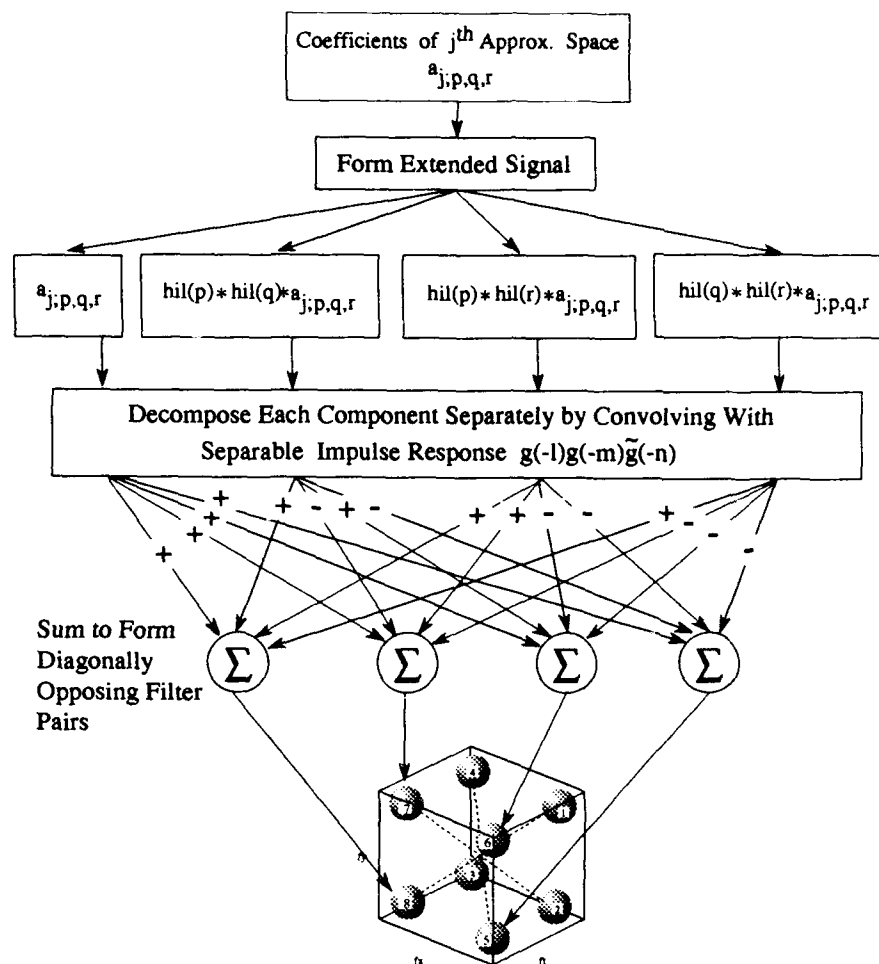


Figure 47. Flow diagram depicting one step in a process designed to capture frequency information contained in diagonally opposing frequency regions of a symmetric wavelet filter. The j^{th} approximation coefficients are decomposed into four sets of d^T detail coefficients which are then summed to extract the diagonally opposing frequency information.

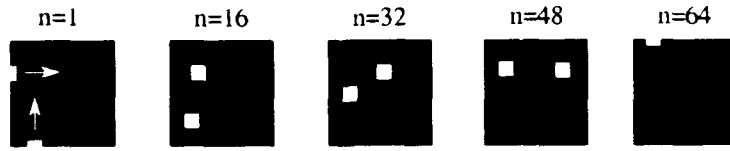


Figure 48. Several frames of 64×64 imagery containing two identical rectangles traveling with the velocity components $(v_x = 1, v_y = 0)$ frame/sec. and $(v_x = 0, v_y = 1)$ frame/sec.

components of the moving object. For example, consider the case of an object that is moving either purely horizontally or purely vertically. The plane containing its Fourier transform will then consist of one unknown velocity component, v_x or v_y , as given by:

Horizontal Motion: $f_t = -v_x f_x$

Vertical Motion: $f_t = -v_y f_y$

In this case, a single directionally selective wavelet filter with a center frequency pair f_t, f_x or f_t, f_y , that matches the spectrum of the moving object can unambiguously segment the object from the scene. In essence, a single wavelet filter, then, provides the ability to solve one equation in one unknown. In order to demonstrate this capability, consider the image sequence shown in Figure 48. The image sequence contains two objects traveling at the same speed, however one object moves horizontally while the other moves vertically. Both sequences contain 64 frames of 64×64 grayscale imagery. Now consider the Fourier transforms of the stationary and moving objects.

Figure 49a) displays the magnitude of the 2D Fourier transform of both identical rectangles. This figure was obtained by Fourier transforming a single frame from the image sequence. The regions of support of the horizontal, vertical and diagonal filters generated by two levels of decomposition in a 2D spatial wavelet multiresolution analysis are overlaid in white on the 2D spectrum. The figure in part b) provides a visualization of the planes in spatio-temporal Fourier space containing the 2D Fourier transforms. The spheres in b) represent the symmetric detail filter created by decomposing the image sequence one level in space and time in a motion-oriented multiresolution wavelet analysis. The solid black and gray spheres represent diagonally opposing detail filters that capture objects traveling horizontally (black) or vertically (gray).

In order to segment the two objects, all four discrete signal components in Equation 148 were constructed using the methods described earlier. Each signal component was then decomposed one

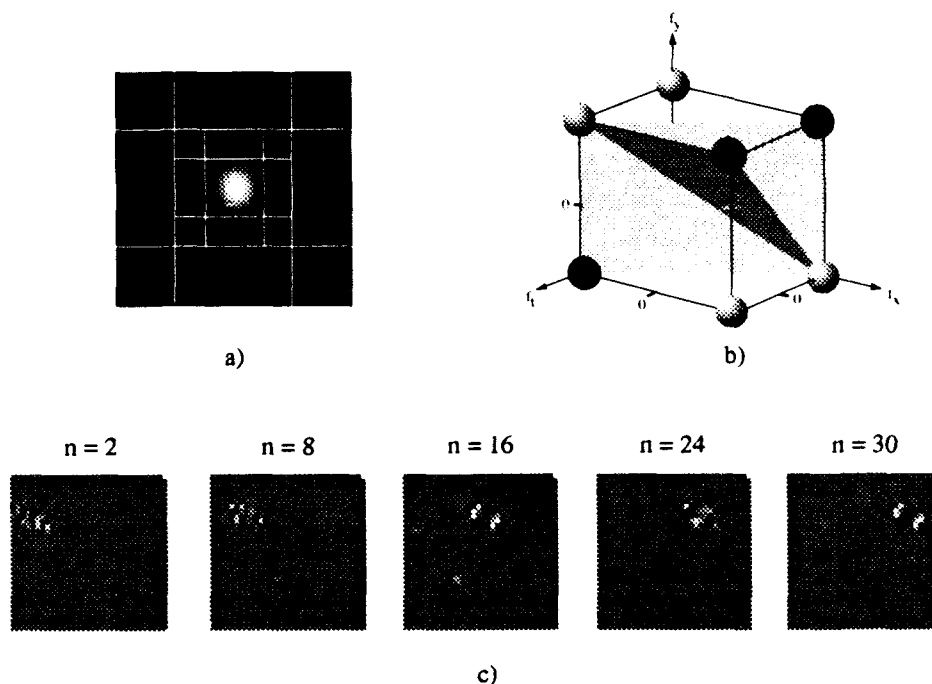


Figure 49. a) 2D FFT of the rectangles in Figure 48. The frequency supports of the spatial wavelet filters generated by two steps in a wavelet multiresolution analysis are overlayed on the FFT. b) A visualization of the planes in Fourier space containing the FFTs of the moving objects. The spheres represent the symmetric wavelet filter generated by one spatial and temporal decomposition level in a motion-oriented multiresolution wavelet analysis. The solid black spheres are selective for horizontal motion while the gray spheres (one is hidden) select for vertical motion. c) The coefficients obtained by segmenting the horizontally moving object using the directionally selective motion-oriented multiresolution wavelet analysis. The filter pair created during this process is highlighted in black in part b).

level in space and time using the motion-oriented multiresolution wavelet analysis. The coefficients for each signal component associated with the symmetric detail filter shown in Figure 49b) were then summed in accordance with the sign conventions in Equations 138 and 139 to produce the filters represented by the diagonally opposing black and gray spheres in part b). The resulting coefficient sequence associated with the filter pair that captures horizontally moving objects is shown in Figure 49c).

This test demonstrates that the directionally selective wavelet filters formed by appropriately summing the decomposition coefficients of the four extended signal components are able to segment two rectangular objects traveling horizontally and vertically with the same speed. However, the test also shows that the coefficients produced by this process are clearly distorted compared to those produced by the symmetric filter in Section 4.5. This distortion is partly attributable to the directionally selective filters which only capture the spatial frequency components in diagonally opposing quadrants of the 2D Fourier transform in Figure 49a). Additionally, some distortion is introduced by the digital Hilbert transform process and by machine precision limitations that affect the coefficient summation process. However, these distortion effects notwithstanding, there is yet a more fundamental problem associated with the segmentation of moving objects using a single diagonally opposing filter pair.

The moving objects in the foregoing example travel in either a purely horizontal or purely vertical direction. Consequently, they each have only one velocity component and their Fourier transforms lie on planes that are described by a single equation in one unknown velocity component (the other component equals zero). Only one diagonally opposing wavelet filter with a known center frequency pair, say (f_{xo}, f_{to}) in the case of a horizontally moving object, is therefore required to solve the plane equation, $f_t = -(f_x v_x + f_y \cdot 0)$, for the unknown velocity component v_x . The problem, of course, is that objects don't typically move in a purely horizontal or vertical direction, so that their corresponding planes in Fourier space are generally described by one equation in two unknown velocity components, i.e., $f_t = -(f_x v_x + f_y v_y)$. This implies objects moving in many different directions can produce spectra that lie in the passband of a single diagonally opposing filter. Consequently, a single filter cannot unambiguously segment objects moving in arbitrary directions. This problem is resolved in the following section by computing the response of multiple filters at a given location in the scene and combining the responses to solve for both unknown velocity components.

5.2.4 Computing Optical Flow. The previous sections showed that the symmetric wavelet filter produced by the motion-oriented multiresolution wavelet analysis acts a scalar motion sensor in that it responds to object speed (a scalar quantity) rather than object velocity (a vector quantity consisting of speed *and* direction). Consequently, a method was presented which combined the properties of the Hilbert transform with the motion-oriented multiresolution wavelet analysis to produce a bank of directionally selective wavelet filters. It was shown, however, that a single directionally selective wavelet filter is not sufficient to unambiguously determine the direction of a moving object. The purpose of this section, therefore, is to present a method that allows one to unambiguously compute direction of motion by combining the responses of two directionally selective wavelet filters. Specifically, the method yields the x and y velocity components - or optical flow - of a moving brightness pattern at each point in the image plane.

5.2.4.1 Concept. Consider again the case of a simple 1D object traveling at some speed v_x as previously discussed in Chapter II. Recall that the Fourier transform of the moving object lies on a line in 2D frequency space defined by $f_t = -v_x f_x$. If the largest digital spatial frequency of the stationary object is π radians, then the frequency support of the moving object is depicted by the line in Figure 50. Here, the line is superimposed on the wavelet filters generated by multiple spatial and temporal decompositions in a 2D motion-oriented multiresolution wavelet analysis as discussed in Chapter IV. The shaded regions represent the wavelet detail filters produced by one spatial decomposition and three temporal decompositions.

The digital center frequencies of each of the filters in the upper right quadrant of the filter bank are shown along the spatial and temporal frequency axes of the 2D frequency space in Figure 50. The wavelet detail filter in the upper right corner of the quadrant (marked with a "1") is produced by convolving the rows and columns of the sampled spatio-temporal input signal with the discrete QMF g filter. Since the digital center frequency of the discrete 1D filter is π radians, the spatial and temporal digital center frequencies of the 2D filter are also π radians.

Now consider the filter marked with a "2" located in the upper right quadrant of the frequency plane. The digital spatial center frequency of the filter remain π radians; however, since the rows (time axis) of the signal have twice been convolved with the discrete temporal wavelet and decimated by a factor of two, the temporal center frequency is now $\frac{\pi}{2}$ radians. Because the line formed in

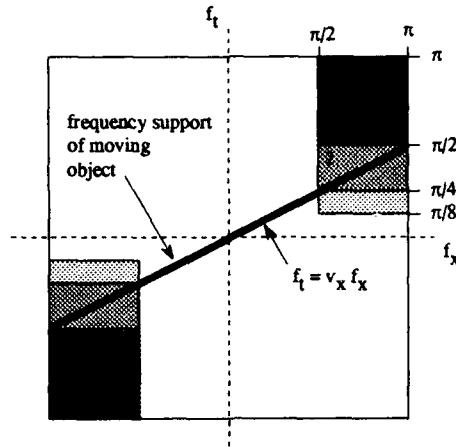


Figure 50. The linear supporting region in 2D frequency space of the Fourier transform of a 1D object translating with some velocity v_x . The line is superimposed on the wavelet filter bank produced by a 2D directionally selective motion-oriented multiresolution wavelet analysis. The shaded regions represent the diagonally opposing wavelet detail filters produced by one spatial decomposition and three temporal decompositions. The digital center frequencies of the filters are show in the upper right quadrant.

frequency space by the Fourier transform of the moving object lies through this filter, its (f_x, f_y) digital center frequency pair $(\frac{\pi}{2}, \pi)$ can be used to estimate the velocity of the object at each point along the subsampled spatial axis as follows:

$$\begin{aligned}
 v_x &= -\frac{f_t}{f_x} \\
 &= -\frac{\pi}{2\pi} \\
 &= -\frac{1}{2} \text{ frame/sec}
 \end{aligned} \tag{149}$$

In order to extend the 2D velocity filtering concept to three dimensions, next consider Figure 51a). This figure shows several diagonally opposing filter pairs obtained by decomposing an "extended" signal once in space and twice in time and summing the coefficients following the sign convention in Equation 134. The digital center frequencies of the horizontal, vertical and diagonal filter pairs are identified by their respective digital center frequency triplets.

Now suppose a single object is traveling diagonally from the lower left to upper right across an image plane. The plane in 3D frequency space containing the object's Fourier transform is depicted in

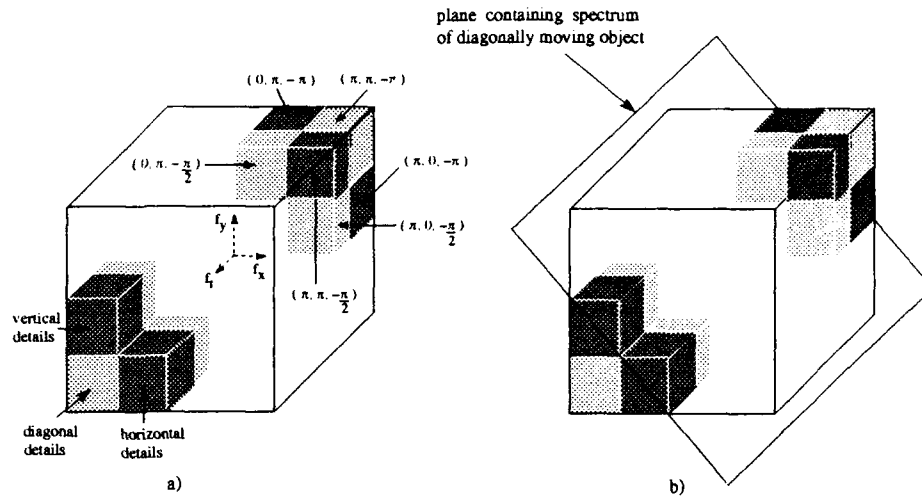


Figure 51. a) The digital center frequency triplets (f_x, f_y, f_t) of the diagonally opposing filters generated by decomposing a discrete input signal one level in space and two levels in time using the directionally selective motion-oriented multiresolution wavelet analysis. b) A visualization of the plane in 3D frequency space containing the Fourier transform of a 2D brightness pattern moving diagonally from the lower left to the upper right hand corner of an image plane.

Figure 51b). Notice that the plane slices through the two darkly shaded filters with center frequencies $(0, \pi, -\pi)$ and $(\pi, 0, -\pi)$. Both center frequency pairs can be used to estimate the velocity components of the moving object by solving the 2×2 system of equations given by

$$\begin{aligned} -\pi &= -(0 \cdot v_x + \pi \cdot v_y) \\ -\pi &= -(\pi \cdot v_x + 0 \cdot v_y) \end{aligned} \quad (150)$$

Solving these equations then yields the velocity components $v_x = v_y = 1$ frame/sec.

Clearly, the accuracy of the velocity estimate depends on the spatio-temporal frequency characteristics of the analyzing filter. A filter with a sharp transition region (i.e., narrow variance) will provide a more accurate estimate of the true velocity and it will reduce the amount of overlap, or inter-band aliasing, between neighboring filters. This also reduces the susceptibility of the velocity estimate to noise. On the other hand, increasing the frequency resolution of the filter forces one to pay the price of reduced resolution in space and time. Here, the non-homogeneous multiresolution wavelet analysis developed under this research effort offers a distinct advantage over the conventional homogeneous

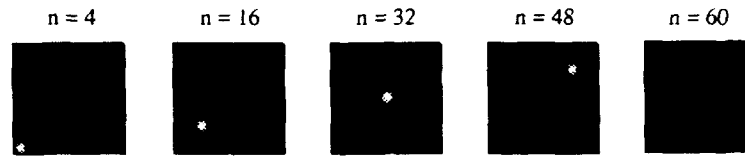


Figure 52. Several frames of 64×64 imagery containing a gaussian brightness pattern traveling with the velocity components ($v_x = v_y = 1$) frame/sec. The variance of the gaussian was chosen to prevent spatial and temporal aliasing.

discrete wavelet transform algorithm in that it allows one to minimize the effects of the resolution trade-off by constructing the 2D analyzing wavelet filter from 1D filters of different orders.

For example, suppose several objects with significantly different spatial frequency characteristics are traveling at similar speeds along the 1D spatial axis. The temporal frequency resolution requirements are more stringent than those for the spatial frequencies. Thus, one might design a more computationally efficient separable 3D wavelet filter by combining a higher order temporal wavelet with a lower order spatial wavelet. The homogeneous discrete multiresolution wavelet analysis, on the other hand, would require one to use the higher order, more computationally expensive, filter in both the spatial and temporal dimensions in order to meet the temporal design criteria.

5.2.4.2 Wavelet Velocity Estimation Algorithm. The previous discussion focused on some general concepts associated with using the directionally selective motion-oriented multiresolution wavelet analysis to compute the velocity of a single 1D and a 2D object moving in a space-time image sequence. This section presents a "multiresolution wavelet velocity estimation algorithm" used to compute the flow field for the more general case of multiple 2D objects moving in a 3D image sequence. In order to familiarize the reader with the details of the 3D algorithm, it is first applied to a relatively simple image sequence containing a single moving 2D brightness pattern. It is then applied to more general scenarios in which multiple objects with different sizes and velocities move in noisy and occluding backgrounds. Several frames of the first test image sequence are shown in Figure 52. The sequence contains 64 frames of 64×64 grayscale imagery. The moving brightness pattern consists of a single gaussian traveling diagonally across the field of view with the velocity components ($v_x = v_y = 1$) frame/sec. The variance of the gaussian was chosen to prevent spatial and temporal aliasing as discussed in Chapter IV.

The first stage of the multiresolution wavelet velocity estimation algorithm employs the discrete directionally selective motion-oriented multiresolution algorithm previously summarized by the flow diagram in Figure 47. The four components of the extended $N \times N \times N$ (row, column, frame) image sequence are first constructed and each is decomposed over all possible resolution levels in space and time. The resulting coefficient sequences associated with each symmetric detail filter in the non-conventional wavelet decomposition are then appropriately summed to extract the information contained in diagonally opposing filter pairs. For example, denote the sampled input image sequence in Figure 52 by $a_{0,0}$, and assume it represents the discrete coefficients obtained by projecting the input signal onto the zeroth approximation space $A_{0,0}f$. As was the case in Chapter IV, the two subscripts separately represent the spatial and temporal resolution level of the projection. Furthermore, assume that the discrete extended signal has been formed as described in the previous section and that the 3D component sequences are given by $c_{0,0}^1, c_{0,0}^2, c_{0,0}^3$ and $c_{0,0}^4$ where $c_{0,0}^1 = a_{0,0}$.

Decomposing the four $N \times N \times N$ extended signal components one level in space then generates four sets of three $\frac{N}{2} \times \frac{N}{2} \times N$ detail coefficient sequences denoted by $d_{-1,0}^{c,1}, d_{-1,0}^{c,2}$, and $d_{-1,0}^{c,3}$ where $c = 1, 2, 3, 4$ stands for "component". Each of the $4 \times 3 = 12$ spatial detail sequences is then decomposed over all possible resolutions in time, while retaining only the temporal detail coefficient sequences as discussed in Chapter IV. If the order of the temporal filter is such that only two temporal decompositions occur before the length of the temporal filter exceeds the number of frames in the decomposed image sequence, then the decomposition process will produce an $\frac{N}{2} \times \frac{N}{2} \times \frac{N}{2}$ and $\frac{N}{2} \times \frac{N}{2} \times \frac{N}{4}$ coefficient sequence for each of the twelve spatial detail coefficient sequences. This process ultimately produces 24 sequences denoted by $d_{-1,t}^{c,f}$ where $f = \text{"filter"} = 1, 2, 3$ and $t = \text{"time"} = -1, -2$. Note that each of the 24 sequences has spatial dimensions of $\frac{N}{2} \times \frac{N}{2}$.

The final step in the first stage of the wavelet velocity estimation algorithm is to sum the detail coefficients to produce four diagonally opposing frequency pairs for each of the three filters at each temporal resolution. This process generates a total of 24 coefficient sequences which can be described by the equations

$$\begin{aligned} d_{r18;-1,t}^f &= \sum_p \sum_q \sum_r d_{-1,t}^{1,f} - d_{-1,t}^{2,f} - d_{-1,t}^{3,f} - d_{-1,t}^{4,f} \\ d_{r27;-1,t}^f &= \sum_p \sum_q \sum_r d_{-1,t}^{1,f} + d_{-1,t}^{2,f} - d_{-1,t}^{3,f} + d_{-1,t}^{4,f} \end{aligned}$$

$$\begin{aligned}
d_{r36;-1,t}^f &= \sum_p \sum_q \sum_r d_{-1,t}^{1,f} - d_{-1,t}^{2,f} + d_{-1,t}^{3,f} + d_{-1,t}^{4,f} \\
d_{r45;-1,t}^f &= \sum_p \sum_q \sum_r d_{-1,t}^{1,f} + d_{-1,t}^{2,f} + d_{-1,t}^{3,f} - d_{-1,t}^{4,f}
\end{aligned} \tag{151}$$

where, again, f is the spatial detail filter, $r18$ thru $r45$ represent the four diagonally opposing frequency pairs, $t = -1, -2$ is the temporal resolution level, $p = q = \frac{N}{2}$, and $r = \frac{N}{2}$ or $\frac{N}{4}$ depending on the temporal resolution level. The sign conventions are those used in Equations 134 through 139.

In the second stage of the algorithm, the coefficients associated with all 24 diagonally opposing detail filters are compared to find the two largest coefficients across all time at each point in the subsampled $\frac{N}{2} \times \frac{N}{2}$ image array. The digital center frequencies of the diagonally opposing detail filters associated with the two largest coefficients at each location are then recorded and passed on to the third stage of the algorithm. As shown previously in Figure 51, the value of each frequency element in the triplet is determined by the filter type (i.e., horizontal, vertical or diagonal), the diagonally opposing passband region, and the number of times the original coefficient sequence was decomposed along the spatial or temporal axes to generate the filter.

The third stage of the wavelet velocity estimation algorithm computes and assigns a velocity vector to each location in the subsampled image array. This is accomplished by solving the plane equation $f_t = -(v_x f_x + v_y f_y)$ at each location using the center frequency triplet pair computed for that location in the second stage of the algorithm. The third stage solves the two-by-two system using Cramer's rule, which, given the center frequency pair (f_{x1}, f_{y1}, f_{t1}) and (f_{x2}, f_{y2}, f_{t2}) , computes the velocity components as follows

$$\begin{aligned}
v_x &= -\frac{(f_{y2}f_{t1} - f_{y1}f_{t2})}{D} \\
v_y &= -\frac{(f_{x1}f_{t2} - f_{x2}f_{t1})}{D}
\end{aligned} \tag{152}$$

where $D = f_{x1}f_{y2} - f_{y1}f_{x2}$ is the determinant of the two-by-two system. In the event that the determinant at a particular location equals zero, the velocity components at that point are both set to zero. Figure 53 shows the output of the wavelet velocity estimation algorithm at the first spatial decomposition level for the input image sequence shown previously in Figure 52. The flow map was obtained by passing the 32×32 subsampled array of velocity components computed by the third stage of the wavelet velocity estimation algorithm to the MATLAB[®] flow generation package called

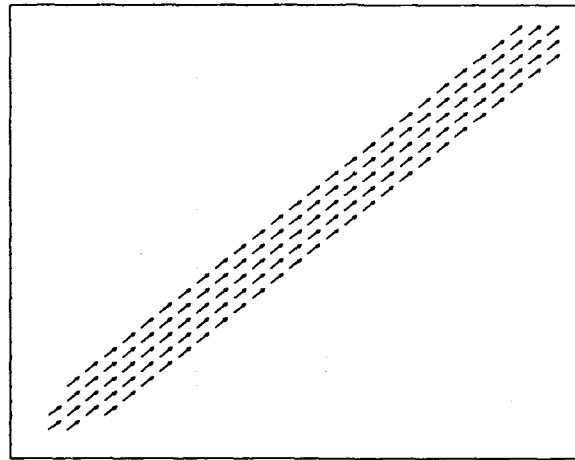


Figure 53. The optical flow map produced by decomposing the image sequence in Figure 52 one level in space and multiple levels in time. The vectors indicate the direction and speed of the moving object.

“quiver.” The length and direction of the velocity vectors indicate the speed and direction of the moving object. The image sequence was decomposed using a truncated cubic spline (23 taps) in space and a Daubechies’ 12 in time. As discussed in Chapter II, a cubic spline was chosen for the spatial component of the non-homogeneous 3D wavelet filter to avoid introducing phase distortions associated with asymmetric Daubechies’ filters.

5.2.4.3 Results. The flow map in Figure 53 demonstrates the ability of the wavelet velocity estimation algorithm to characterize motion in a simple scenario consisting of a single object moving against a black background. This section presents the results obtained by applying the algorithm to more complicated motion scenarios. In the first test, the scenario is complicated by the addition of a second object traveling in the same direction but at a different speed (Figure 54a). The velocity of the second object is ($v_x = v_y = .5$) frames/sec. Figure 54b) shows the flow map obtained by decomposing the sequence one level in space using a 3D wavelet filter constructed from a 23 tap truncated cubic spline in space and a Daubechies 4 in time. A smaller order temporal filter was used in this example in order to generate more decomposition levels in time. This in turn increases the number of “speed bins” in the directionally selective motion-oriented multiresolution wavelet filter bank, thereby enhancing

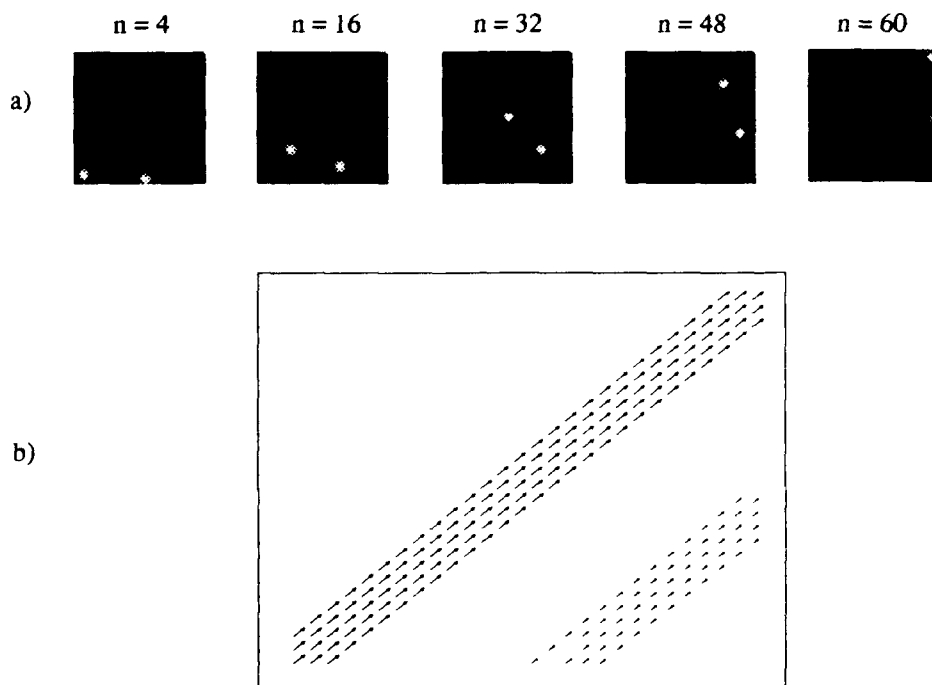


Figure 54. a) Several frames of 64×64 imagery containing two identical gaussian brightness patterns traveling in the same direction but at different speeds. The velocities of the two objects are $(v_x = v_y = 1)$ frames/sec. and $(v_x = v_y = .5)$ frames/sec. b) The optical flow map produced by decomposing the image sequence one level in space using a non-homogeneous 3D wavelet filter constructed from a spatial cubic spline truncated to 23 taps and a Daubechies 4 temporal filter.

the velocity estimation capabilities of the system. The flow map clearly shows that the wavelet-based motion analysis method is able to discriminate between both objects by employing the space/time - frequency localization properties of the multiresolution wavelet analysis. The next test sequence shows that the method is also able to correctly compute the optical flow of multiple objects traveling in several different directions at different speeds.

Figure 55a) shows several frames of a $64 \times 64 \times 64$ image sequence containing four objects traveling with the (v_x, v_y) velocity components $(1, 1)$, $(-1, 1)$, $(1, 0)$ and $(0, 1)$ frames/sec. The image sequence was designed to test the algorithm's response to discontinuities in the velocity flow field. These phenomena occur where the paths of the moving objects cross, and where the vertically moving object in the lower right hand corner abruptly vanishes halfway across the field of view (referred

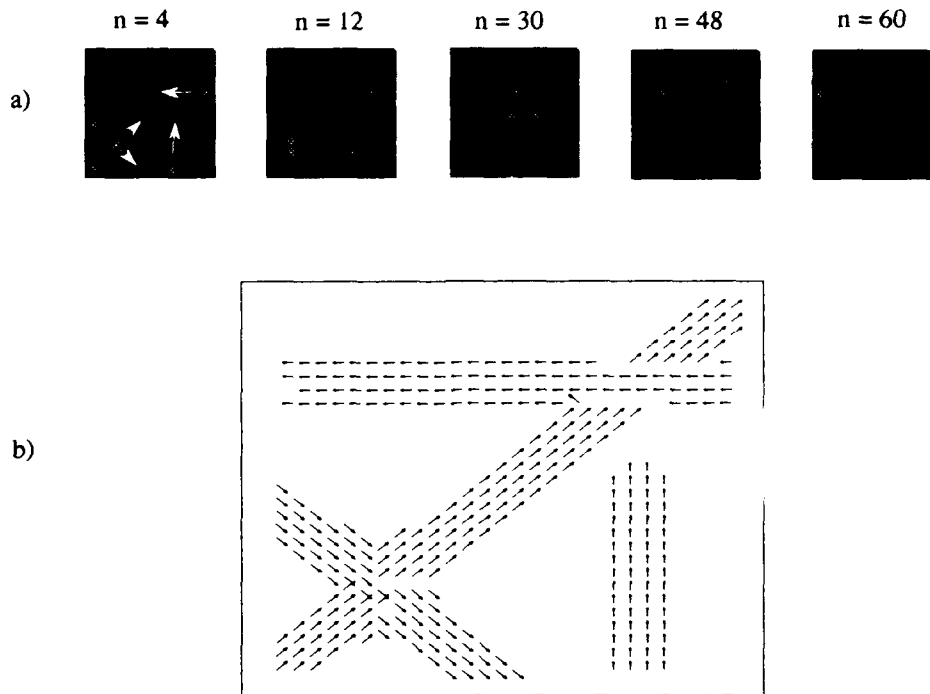


Figure 55. a) Several frames of 64×64 imagery containing four identical gaussian brightness patterns traveling at different speeds and in different directions. b) The optical flow map produced by decomposing the image sequence one level in space using a non-homogeneous 3D wavelet filter constructed from a spatial cubic spline truncated to 23 taps and a Daubechies 12 temporal filter.

to as a motion “sink”). Part b) displays the flow map obtained by decomposing the image sequence one level in space using a 3D wavelet filter constructed from a 23 tap truncated cubic spline in space and a Daubechies 12 in time. From a qualitative perspective, the flow vectors at the points in the scene where the moving objects cross paths and where the vertically moving object abruptly stops, do not seem appreciably affected by the discontinuities in the velocity field. The velocity vectors at the locations where the objects cross paths generally match the direction and speed of at least one of the overlapping objects. Additionally, the flow vectors at the motion sink do not display the random behavior associated with more common spatio-temporal gradient flow computation techniques (see Chapter II). Both these phenomena are attributable to the “flow averaging” effects of the spatial and temporal wavelet convolution processes. The next example demonstrates the multiresolution properties of the wavelet velocity estimation algorithm.

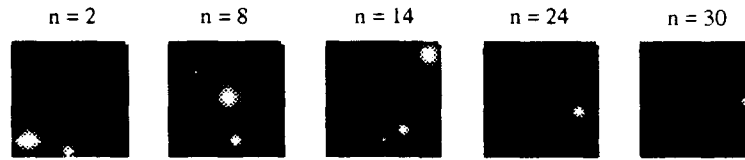


Figure 56. Several frames of 32×32 imagery containing two gaussian brightness patterns with different variances and velocities. The large object travels at four times the speed of the smaller object.

After completing the third stage of the flow computation process, the wavelet velocity estimation algorithm returns control to the first stage. Here, the $\frac{N}{2} \times \frac{N}{2} \times N$ spatial approximation coefficient tensors generated by the first spatial decomposition are again decomposed one level in space and multiple levels in time to yield an optical flow map with the spatial dimensions $\frac{N}{4} \times \frac{N}{4}$. The entire process then recursively repeats until a spatial decomposition level is reached where the number of taps in the spatial component of the separable quadrature mirror filter exceeds the number of approximation coefficients along one spatial dimension of the subsampled image tensor. Since each spatio-temporal decomposition level corresponds to a particular spatial scale (size) and object speed, the wavelet multiresolution velocity estimation algorithm provides the ability to discriminate between objects with different sizes traveling at different velocities. This capability is demonstrated by the next test image sequence shown in Figure 56. The sizes and velocities of the gaussians were chosen to demonstrate the ability of the algorithm to discriminate between large objects traveling fast and small objects traveling slow. Recall that the "conventional" $L_2(\mathbb{R}^3)$ multiresolution analysis developed in Chapter III did not provide this capability.

The results of the multiresolution test are shown in Figure 57. In order to clearly show the size and velocity discrimination properties of the algorithm, a separate flow map was generated at each spatial and temporal decomposition level. An 11 tap cubic spline spatial filter and a 12 tap Daubechies temporal filter were used to construct the separable 3D wavelet filter; therefore, the velocity estimation algorithm allows two spatial and two temporal decompositions. This yields four multiresolution flow maps, where each map captures a different size and speed combination for the gaussian brightness patterns used in the test. The test results, which contain a single flow field at the spatial and temporal resolutions associated with large, fast and small, slow objects, clearly demonstrate the size and velocity resolution properties of the wavelet multiresolution velocity estimation algorithm.

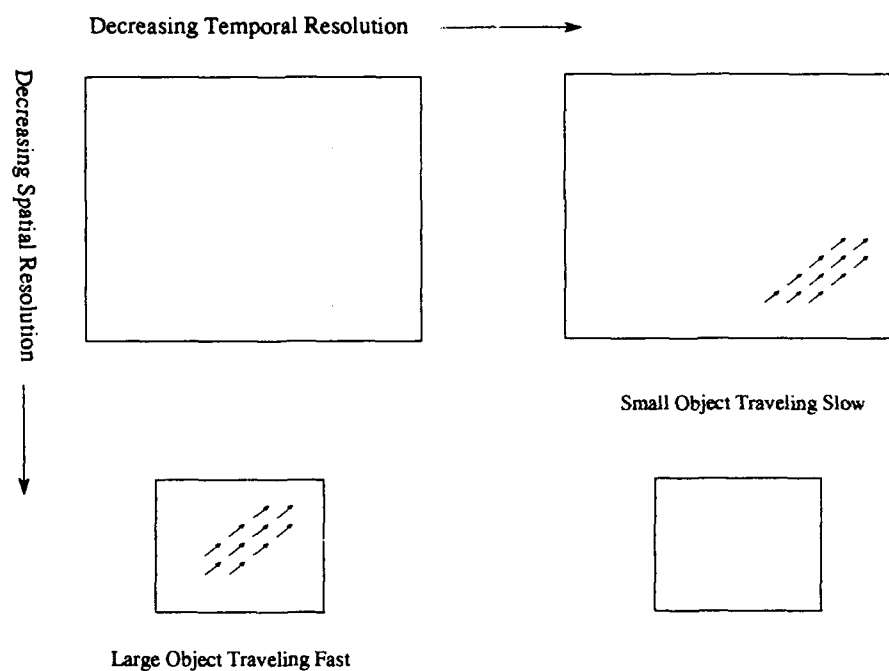


Figure 57. Four multiresolution optical flow maps generated by computing the optical flow of the image sequence in Figure 56 at multiple resolutions in space and time. The results show that the wavelet multiresolution velocity estimation algorithm is able to clearly discriminate between large objects traveling fast, and small objects traveling slow.

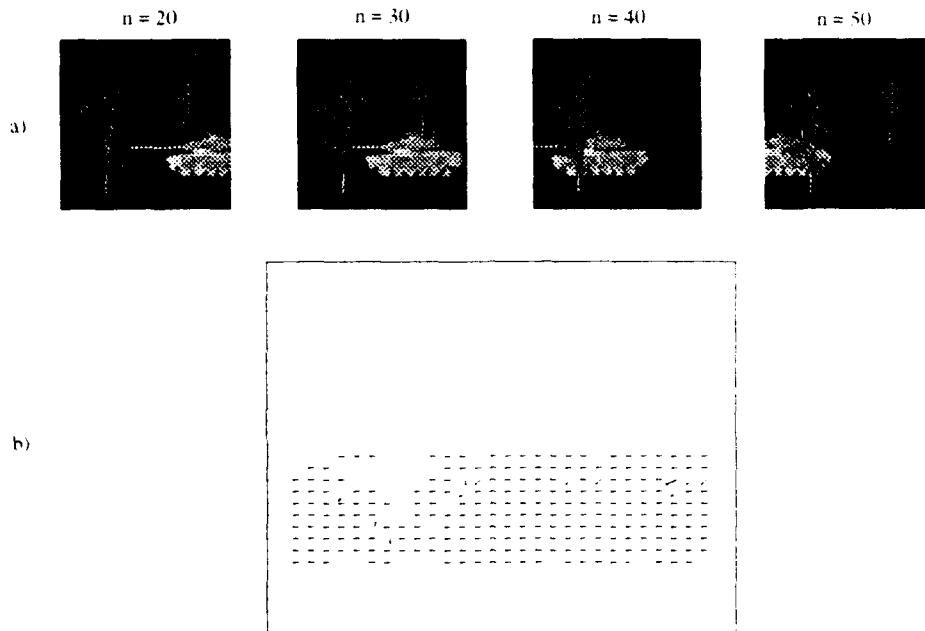


Figure 58. a) A simulated tank moving across an occluded field of view. The simulated illumination source located to the left and behind the reader causes the tank's pixels to scintillate as the tank changes position with respect to the source. Additionally, the tank passes in front of the tree on the right and behind the tree on the left. b) Optical flow field generated by the wavelet multiresolution velocity estimation algorithm.

The final test conducted in this section applies the velocity estimation algorithm to a more realistic image sequence containing a tank moving across an occluded field of view (Figure 58a)). The $64 \times 64 \times 64$ image sequence was constructed on a Silicon Graphics computer using the computer aided software design package BRL-CAD[®]. In this image sequence, the tank moves across the field of view at a constant velocity, passing in front of the tree on the right and partially behind the tree on the left. Reflections from a simulated illumination source located behind and to the left of the reader cause the pixels on the tank's surface to scintillate as it changes position with respect to the source. The image sequence was decomposed using a 23 tap cubic spline in space and a Daubechies 12 in time.

The velocity flow field generated by applying the wavelet multiresolution velocity estimation algorithm to the simulated tank image sequence is shown in Figure 58b). The flow field generally indicates horizontal motion at a constant speed over the regions in the field of view corresponding to

the moving tank. However, there are erroneous flow vectors located at several locations in the scene. These flow aberrations are primarily attributable to pixel scintillations on the tank's surface which violate the constant intensity assumption of the algorithm. The rapidly fluctuating pixels generate temporal frequency components which lie off the object's motion plane, and which create the illusion of motion in non-horizontal directions by exciting the filter pairs associated with these directions. In addition to the scintillating pixels, the velocity discontinuities near the base of the occluding tree also appear to cause minor problems for the velocity estimation algorithm. However, erroneous flow information around these discontinuities, like the motion sink in a previous example, is for the most part eliminated by the spatio-temporal averaging effects of the 3D wavelet correlation. The following section presents a unique competitive-cooperative flow restoration mechanism that employs a gated dipole filter to correct flow aberrations caused by these and other types of noise sources.

5.3 A Cooperative-Competitive Flow Restoration Mechanism

The final example in the previous section demonstrates the effect two common noise sources have on the accuracy of the flow vectors computed by the wavelet multiresolution velocity estimation algorithm. The purpose of this section is to present a unique flow restoration mechanism that finds and corrects localized flow inconsistencies generated by these noise sources. The section begins with a brief discussion of the gated dipole filter as first proposed by S. Grossberg as part of his Boundary Contour System (22). Next, a methodology is presented that combines a modified version of Grossberg's gated dipole filter with a cooperative-competitive strategy that rewards and enhances consistent flow behavior and removes flow inconsistencies. Several examples are then provided which demonstrate the flow correction capabilities of this methodology.

5.3.1 Modified Gated Dipole Filter. A gated dipole filter is a non-linear interpolation filter that was first developed by S. Grossberg for the purpose of "filling in" partially completed object boundaries (22). This section describes a modified version of Grossberg's dipole filter designed to fill in partially completed flow fields and re-orient flow vectors that lie outside a prescribed orientation bandwidth established by the orientations of neighboring flow vectors. The output of the modified

dipole filter, D_{ijk} , for the k th orientation at the i, j th pixel location is given by

$$D_{ijk} = \sum_{r,p,q} [I_{rpq}] [F_{pqij}^k]^+ + \sum_{r,p,q} [I_{rpq}] [G_{pqij}^k]^+ \quad (153)$$

where I denotes pixel intensity and the $+$ symbol indicates rectification. The variable k represents a particular flow vector orientation, where it is assumed that the flow vector orientations have been discretized into M levels. Furthermore, it is assumed that the "composite" flow map generated by the multiresolution wavelet velocity estimation algorithm has been subdivided into M oriented flow maps, where the oriented flow map M_k contains only the vectors in the composite flow map that lie in the k th direction. The filter is centered at the flow map coordinates i, j and the summation occurs over all coordinates p, q in the " r -neighborhood" of oriented flow maps surrounding the k th orientation.

The functions F_{ijpq}^k and G_{ijpq}^k in Equation 153 are given by

$$\begin{aligned} F_{ijpq}^k &= e^{-2(N_{ijpq}/P-1)^2} [\cos(Q_{ijpq} - k)^T]^+ \\ G_{ijpq}^k &= -e^{-2(N_{ijpq}/P-1)^2} [\cos(Q_{ijpq} - k)^T]^+ \\ N_{ijpq} &= \sqrt{(i-p)^2 + (j-q)^2} \\ Q_{ijpq} &= \arctan\left(\frac{j-q}{i-p}\right) \end{aligned} \quad (154)$$

where each function represents one lobe of a dipole receptive field. Figure 59 shows both lobes plotted on a "horizontal" orientation plane. The activation of both lobes is plotted as a three dimensional surface above the plane.

Now consider one lobe of the dipole plotted as a 2D projection onto the k th orientation plane as shown in Figure 60. The major parameters that determine the characteristics of the dipole lobe are N_{ijpq} , Q_{ijpq} , P , and T . N_{ijpq} computes the distance from the dipole center (i, j) to a surrounding point (p, q) . Q_{ijpq} computes the angle of the line segment joining (i, j) and (p, q) . The major axis of the dipole lobe lies in the k th direction and is represented by the vector $\hat{k}$.

The cosine term, $\cos(Q_{ijpq} - k)$, determines the orientational tuning characteristics of the filter by measuring how parallel the line segment $(i, j \rightarrow p, q)$ is to the vector $\hat{k}$. The maximum value of the kernel occurs when $\hat{k}$ and the line segment $(i, j \rightarrow p, q)$ both lie in the same direction. Since the

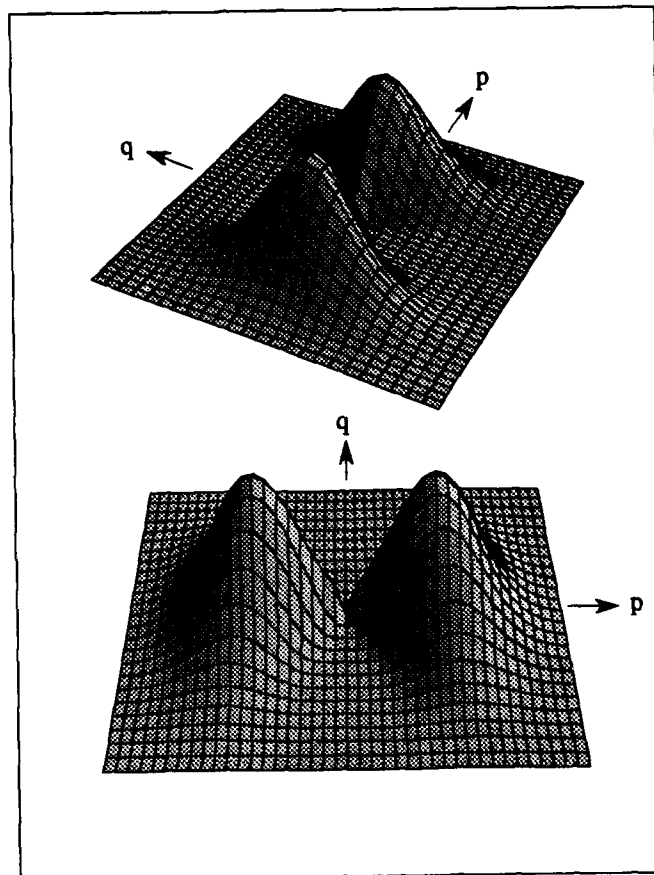


Figure 59. Two views of the right and left hand lobes of a dipole filter lying on a horizontal orientation plane. The lobe activations are represented as 3D surfaces above the orientation plane.

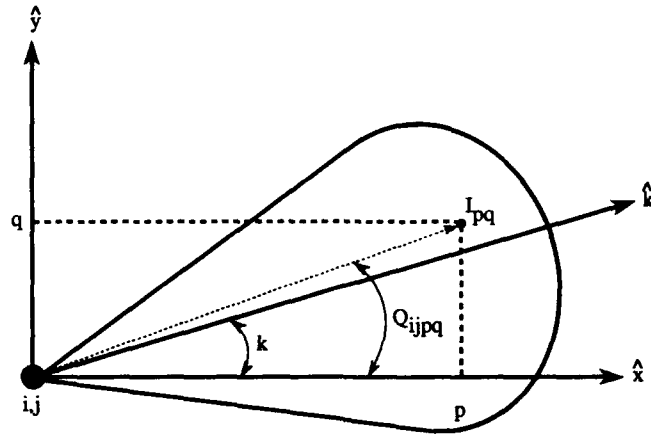


Figure 60. A 2D projection of a dipole lobe onto the k th orientation plane. The major axis of the dipole is oriented in the $\hat{k}$ direction.

cosine term is half-wave rectified, the kernel equals zero at all points (p, q) for which the absolute value of the angle Q_{ijpq} is greater than 90° .

The constant P in the leading exponential term determines the location along the major axis of the highest activation region of the dipole lobe. Figure 61 (top) illustrates how the position of the activation region changes with P . The constant T is an odd integer that determines the sharpness of the orientational tuning by controlling the rate at which the activation region decays on either side of the major axis. Figure 61 (bottom) shows tuning sharpness increases (i.e, the width of F_{ijpq} decreases) with T . Thus, as T decreases, the gated dipole filter is able to fill in missing flow vectors across like oriented, but disjoint vectors in space. This behavior is shown in Figure 62.

As noted previously, the summation parameter r is an element of a small neighborhood of orientations surrounding the k th oriented flow map. The size of the neighborhood determines the "orientation bandwidth" of the gated dipole filter. For example, consider the localized flow field shown in Figure 63a). Although the majority of the flow vectors indicate purely horizontal motion, a small number indicate two slightly different motion directions, presumably caused by noise in the image sequence. Figure 63 shows three oriented flow maps centered around the horizontal flow map that contains the three differently oriented vectors. The horizontally oriented gated dipole filter is positioned at the same location in each flow map. The response at this location, which consists of the sum of the

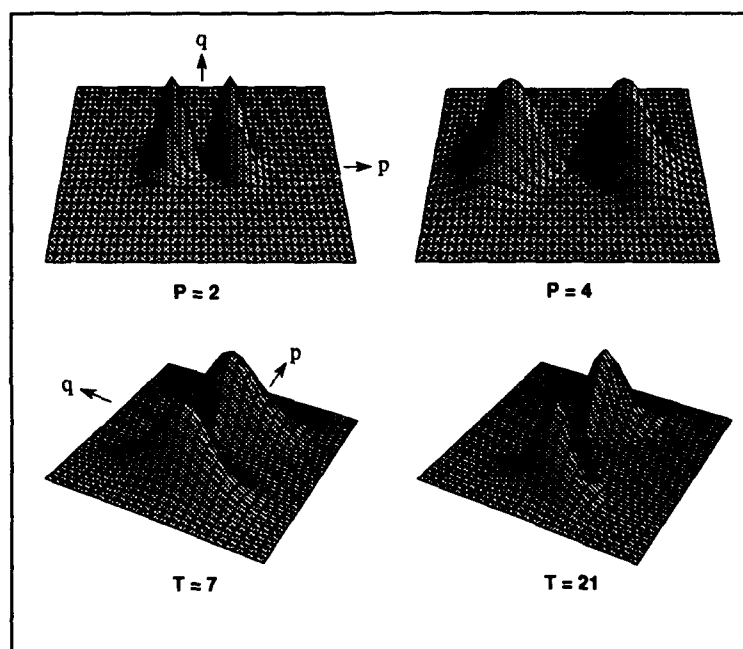


Figure 61. Top: Plots showing how the location of the highest dipole activation region varies with the constant P . Bottom: Plots showing how the decay of the dipole activation region across the major axis (i.e., tuning sharpness) varies with the constant T .

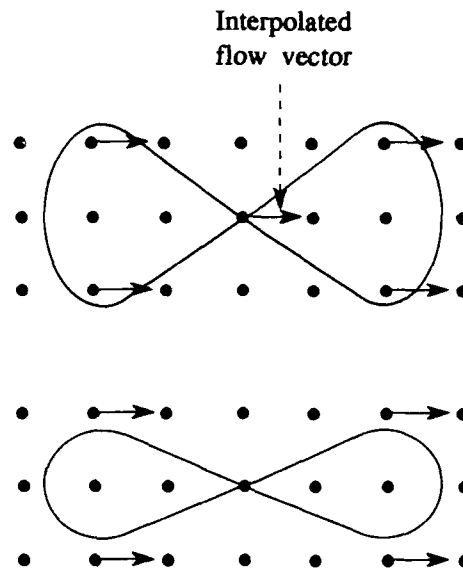


Figure 62. Flow completion properties of dipole filter as a function of T . Top dipole: Small T expands the width of the dipole lobes to complete flow path between two disjoint flow vectors. Bottom dipole: Large T contracts the width of the dipole lobes and prevents the formation of an interpolating flow vector.

responses from each of the three flow maps, is clearly greater than the response of the horizontal flow map alone. Thus, increasing the size of the orientation bandwidth can enhance the overall response of an oriented gated dipole filter. Assuming the horizontally oriented response in this example exceeds the response of all other directionally oriented filters, the flow restoration methodology described in the next section then replaces the distorted vector with the correct, horizontally oriented flow vector, thereby correcting the noise induced flow distortion.

5.3.2 Methodology. The flow restoration methodology presented here employs a cooperative-competitive strategy that reinforces consistent flow behavior and eliminates flow inconsistencies. Consistent flow behavior is defined as follows:

An ensemble of three or more neighboring flow vectors are "consistent" if they possess the same orientation and magnitude, and if they propagate in a direction parallel to their common orientation.

The fundamental premise here is that the path taken by an object as it travels across the field of view should lie precisely in the same direction as its corresponding flow vectors. At first, this may seem

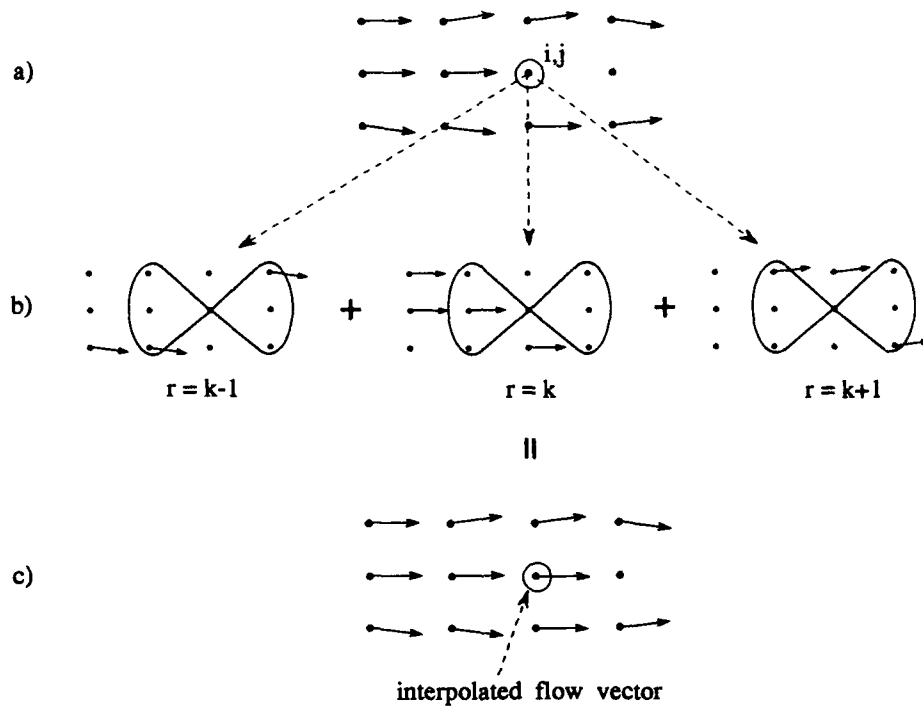


Figure 63. a) Localized flow distortions caused by noise in a 3D image sequence. b) Flow maps contained in the orientation bandwidth surrounding the horizontal flow orientation. A horizontally oriented gated dipole is shown at the position i, j in each of the flow maps. c) Outputs of oriented flow maps sum to form interpolated flow vector at i, j .

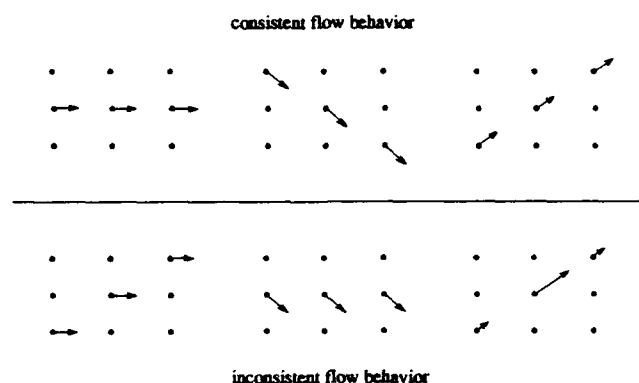


Figure 64. Several examples of consistent and inconsistent flow behavior.

like a trivial assumption, but one only has to examine the flow map of the moving tank in the previous example to find an example of an ensemble of vectors that collectively propagate in one direction (horizontal) yet point in a different direction. Several examples of consistent and inconsistent flow behavior are shown in Figure 64. Incidentally, under this definition, the consistency of a single flow vector or of two neighboring vectors is undetermined.

The cooperative-competitive flow restoration methodology consists of two main stages as shown in Figure 65. The first stage, called the cooperative stage, is designed to reinforce consistent flow behavior. This stage begins by discretizing the orientations of the flow vectors in each of the multiresolution flow maps produced by the wavelet multiresolution wavelet velocity estimation algorithm. For this research, the vector orientations were discretized to positive integer multiples of 22.5° . This generates 16 possible discrete flow orientations between the angles of 0 and 360 degrees. In the next step of the cooperative stage, the composite multiresolution flow maps are decomposed into a set of *oriented* flow maps, where each oriented flow map contains only the vectors in the composite flow map that match one of the discrete orientation levels. In this case, 16 oriented flow maps were generated for each composite multiresolution flow map. Next, each oriented flow map is filtered with an *identically* oriented gated dipole filter to cooperatively reinforce consistent flow behavior in the composite flow map.

In order to determine flow consistency at a given orientation, a temporary, binary oriented flow map is first created in which the existence of a flow vector at a given location is signified by placing a value of 1 at that location. The response of each lobe in a like-oriented gated dipole filter is then

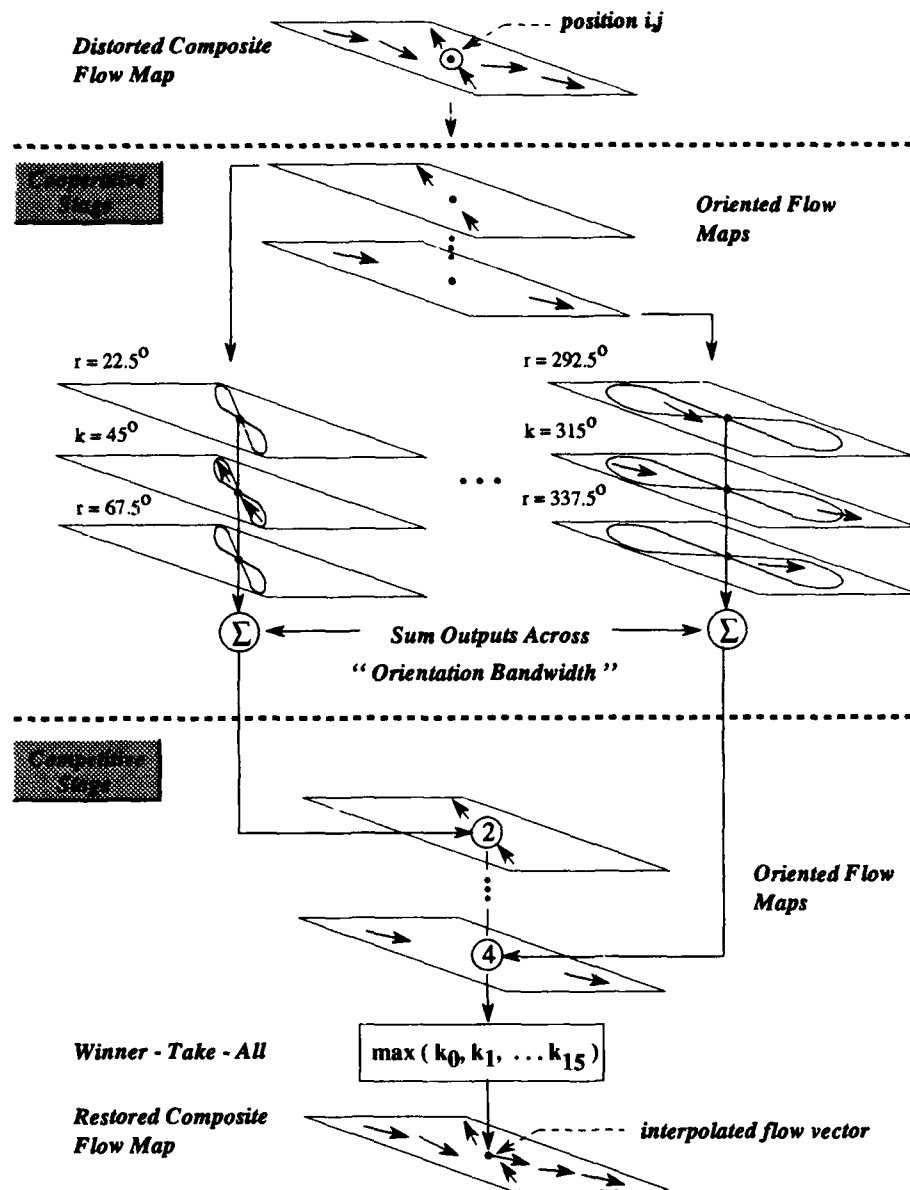
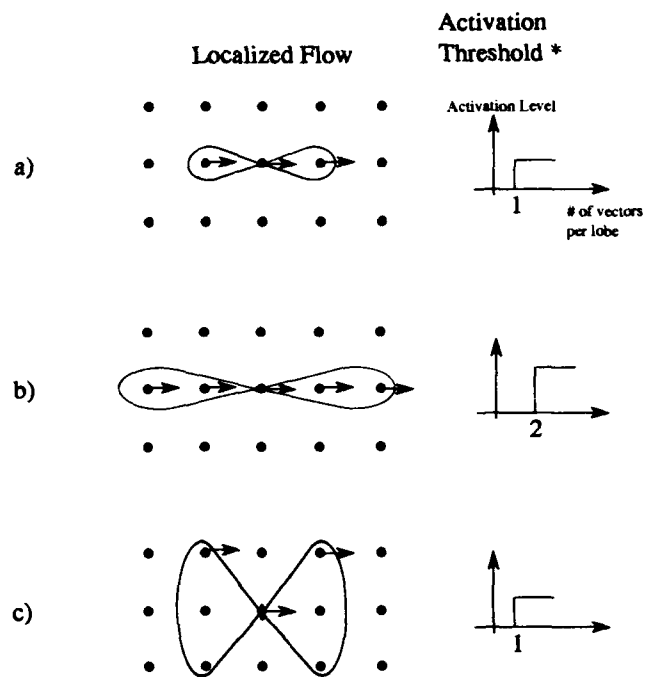


Figure 65. Diagram showing the key elements of the cooperative-competitive flow restoration methodology. Key terms are italicized.

measured at each i, j location in the binary oriented map and in the adjacent binary maps contained in the orientation bandwidth of the gated dipole filter. If the magnitude of the responses in each lobe at a particular location both exceed a pre-defined activation threshold, then their "cooperative" action causes the gated dipole to "fire" and the value of the response magnitude is inserted at the corresponding location in the oriented flow map. This value is eventually converted back to a vector direction when the oriented flow maps are recombined into a single composite flow map. The requirement that both lobes exceed an activation threshold ensures that the flow propagates consistently over a minimum area as determined by the locations of the highest activation regions and the widths of the dipole lobes.

The location of the highest activation region in a dipole lobe is specified by the value of the P parameter in Equation 153. Here, P was chosen so that the locations of the highest activation regions in each lobe occur at the nearest points on opposite sides of center point of the filter as determined by the orientation of the filter (see Figure 66a). Because only three adjacent points are involved in the computation, this setting corresponds to the minimum distance required to achieve flow consistency. One can strengthen the consistency requirement by extending the location of the highest activation region further from the filter center and increasing the activation threshold as shown in Figure 66b). In addition to the location of the highest activation regions, the widths of the receptive fields associated with each lobe also play a significant role in determining flow consistency. Recall that the width of the receptive field is determined by the parameter T in Equation 153. As T decreases, the receptive field of the lobe widens, and more vectors surrounding the main path of the object are included in the flow consistency computation. As shown in Figure 66c), this essentially weakens the flow consistency requirement.

The cooperative stage of the flow restoration methodology ends when the magnitude of the gated dipole filter response is measured and recorded at each location in every oriented flow map. The orientation and magnitude of the restored flow vector is then determined in the second, or competitive, stage of the methodology. In this stage, the response magnitudes at a given location compete across all possible flow orientation levels. Under a winner-take-all rule, the composite flow vector is assigned the orientation of the oriented flow map that possesses the largest response at that location. This winner-take-all strategy ensures that a single flow vector is ultimately assigned to each location in the final composite flow map. The magnitudes of the remaining composite flow vectors are determined by computing the average magnitude of the non-zero flow vectors contained under the gated dipole filter



* Indicates # of flow vectors required to activate 1 lobe of the dipole filter

Figure 66. a) Minimum distance required to achieve flow consistency. The activation threshold for each lobe of the gated dipole filter is shown on the right. b) Strengthening the flow consistency requirement by shifting the location of the highest activation region away from the dipole center and increasing the activation threshold. b) Weakening the consistency requirement by expanding the lobes of the dipole receptive fields.

in the winning oriented flow map. Several examples showing the flow restoration capabilities of the two stage methodology are presented next.

5.3.3 Applications and Results. One of the most challenging problems associated with any optical flow computation is the discrimination of moving targets in the presence of noise. This problem is particularly troublesome for differential flow computation techniques, which generally require continuous spatial and temporal intensity distributions (6). Spatio-temporal integration techniques, on the other hand, may be better suited for this task due to the spatial and temporal averaging effects inherent in the integration process. In particular, this section provides several examples which demonstrate the ability of the integration-based flow computation and correction algorithm developed under this research to discriminate targets obscured by both system and physical noise phenomena.

The first set of examples test the algorithm's ability to discriminate targets moving in an image sequence corrupted by system noise. It is assumed that system noise is an additive process consisting of contributions from several different noise sources including the system's sensor and its electronics package. Under the Central Limit Theorem, the cumulative effects of system noise on each pixel of an output image can be approximated by a gaussian probability density function (PDF) (59). Furthermore, the degree to which this additive gaussian noise is correlated between pixels is determined by the physical properties of the system. In these examples, correlated gaussian noise was obtained by lowpass filtering a white, or uncorrelated, gaussian noise profile using a purely real gaussian filter (59). The degree of correlation between image pixels was determined by the size of the passband of the lowpass gaussian filter.

The amount of noise energy added to each frame in an image sequence was controlled by the standard deviation of the zero-mean gaussian random variable used to generate the noise distribution of each pixel in the image and the average signal intensity of the frame. The relationship between the standard deviation and the noise energy is given by (17)

$$S/N = 20\log_{10} \frac{\text{average signal intensity of frame}}{\text{standard deviation of pixel}} \quad (155)$$

where S/N is the signal to noise ratio of each pixel of the image. In these examples, Equation 155 was first used to determine the required standard deviation for a desired signal-to-noise ratio. The procedures described in (59) were then followed to generate various degrees of spatially correlated

gaussian noise. The degree of correlation was determined by the cutoff frequency, α , of a circularly symmetric real gaussian filter whose frequency response is given by

$$H(u, v) = e^{-\frac{\alpha^2}{8}(u^2 + v^2)^2} \quad (156)$$

where u and v are spatial frequency components. Under these procedures, the degree of correlation increases with α .

The first example is designed to demonstrate the ability of the flow restoration methodology to remove inconsistent flow information contained in a 3D image sequence. Several frames of the $32 \times 32 \times 32$ image sequence used in the first example are shown in Figure 67a). The sequence consists solely of spatially correlated gaussian noise obtained using a filter cutoff frequency of $\alpha = 1$. Figure 67b) shows the multiresolution flow maps produced by the wavelet multiresolution velocity estimation algorithm using an 11 tap cubic spline in space and a Daubechies 4 in time. Since the flow vectors produced at both spatial resolutions demonstrate little if any flow consistency, the flow restoration methodology should eliminate all vectors from each map. This is clearly the case as shown by the "flow restored" maps in Figure 67c).

In the second example (Figure 68), two image sequences were constructed by adding two differently correlated noise patterns to an image sequence consisting of a gaussian intensity distribution moving at $v_x = v_y = 1$ frames/sec. across a 32×32 field of view. In both examples, the signal-to-noise ratio was $-10dB$ and the average signal intensity value was 14.161. The noise sequence in part a) is the same sequence generated in the first example, while part b) was formed from a noise sequence with a higher degree of correlation as given by $\alpha = 10$. Figure 69 shows the flow maps produced after applying the cooperative-competitive flow restoration methodology to each of the image sequences in Figure 67. In both cases, the technique was able to remove flow inconsistencies caused by the gaussian noise sources while retaining the flow vectors associated with the moving object. The fact that the flow restoration process in part a) was less successful than in b) is caused by the presence of higher spatial frequency components that lie near the band limits of the signal spectrum. These frequency components, which are not so strongly present in b), are captured in the first level spatial decomposition whose resolution matches the resolution of the moving object.

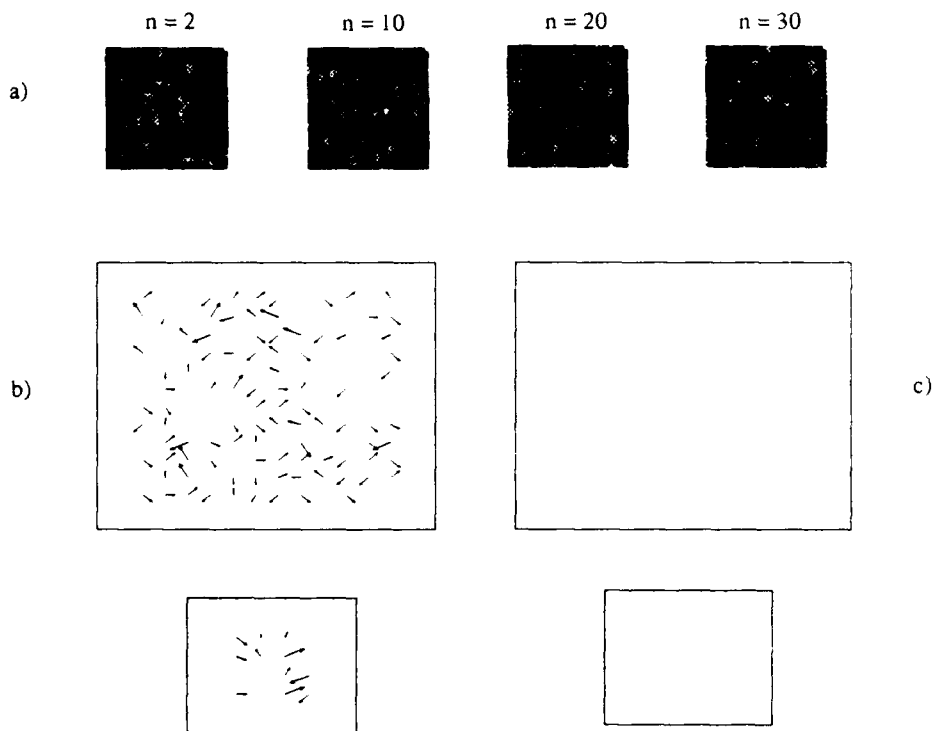
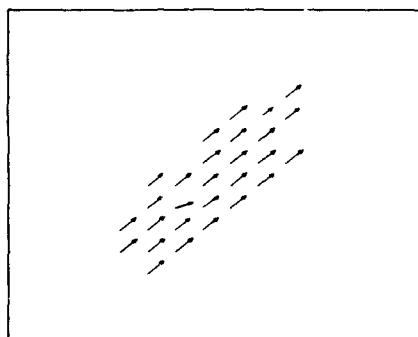


Figure 67. a) Several frames of a $32 \times 32 \times 32$ noise sequence consisting solely of slightly correlated gaussian noise obtained using a filter cutoff frequency of $\alpha = 1$. b) Multiresolution flow maps showing inconsistent flow behavior associated with noise sequence. c) Multiresolution flow maps show flow inconsistencies removed by cooperative-competitive flow restoration methodology.



Figure 68. Several frames from two different image sequences obtained by adding spatially correlated gaussian noise to a $32 \times 32 \times 32$ image sequence containing a gaussian brightness pattern traveling at $v_x = v_y = 1$ frames/sec. The signal-to-noise ratio was $-10dB$ and the average signal intensity value was approximately 14 in both examples. The degree of correlation was obtained from Equation 156 using a) $\alpha = 1$ and b) $\alpha = 10$.

a)



b)

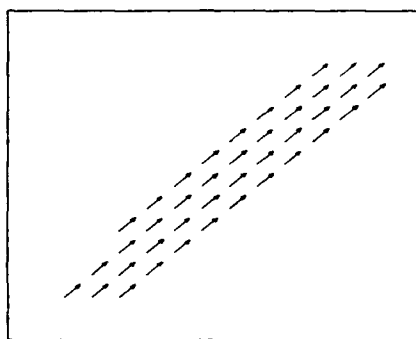


Figure 69. Flow maps produced after applying the cooperative-competitive flow restoration methodology to the image sequences in Figure 67a) and b).

The final example is designed to test the ability of the flow restoration algorithm to remove flow inconsistencies caused by a combination of system and "physical" noise phenomena such as occlusions and reflectance variations caused by movement with respect to a fixed illumination source. In this example, four image sequences were formed by adding four equally correlated noise sequences with different signal-to-noise ratios to the simulated tank image sequence presented in Figure 58 at the end of the previous section. Figure 70 shows the resulting image sequences and their respective signal-to-noise ratios. The cutoff frequency of the filter used to generate the $64 \times 64 \times 64$ equally correlated noise sequences was $\alpha = 2$. The average intensity of the tank sequence (including the trees) was approximately 80. As before, each image sequence was decomposed with a 23 tap cubic spline in space and a Daubechies 12 in time. The flow maps obtained after applying the flow restoration methodology are shown in Figure 71. The flow map at the top of the figure was obtained by applying the wavelet flow computation and restoration algorithm to the "noiseless" image sequence. Notice that flow inconsistencies caused by pixel scintillations on the tank's surface and boundary occlusions have been corrected to indicate the true speed and direction of the tank. The remaining flow maps show the output of the algorithm for the four noisy image sequences in Figure 70. The flow computation and restoration algorithm clearly produced a reasonable approximation of the flow field down to a signal-to-noise ratio of 5. After this point, spurious, noise-induced flow vectors began to appear outside the flow region until, at a signal-to-noise ratio of approximately one, the true flow field became difficult to discern.

5.4 Conclusion

Chapter IV presented a unique, non-homogeneous multiresolution wavelet analysis designed to extract moving objects in a 3D image sequence based on their location, size and speed. It was shown, however, that this "scalar" motion-oriented multiresolution analysis lacked a key attribute of a useful motion analysis tool - directional selectivity. The purpose of this chapter, therefore, was to extend the properties of the motion-oriented wavelet filter bank to form a "vector" motion analysis tool. This new tool is capable of discriminating moving objects based on their location, size, speed *and* direction of movement. The method was based on the formation of four extended real image sequences from a judicious application of 1D spatial and temporal Hilbert transforms. These signals, when decomposed using the discrete motion-oriented multiresolution wavelet analysis developed in Chapter IV, yielded

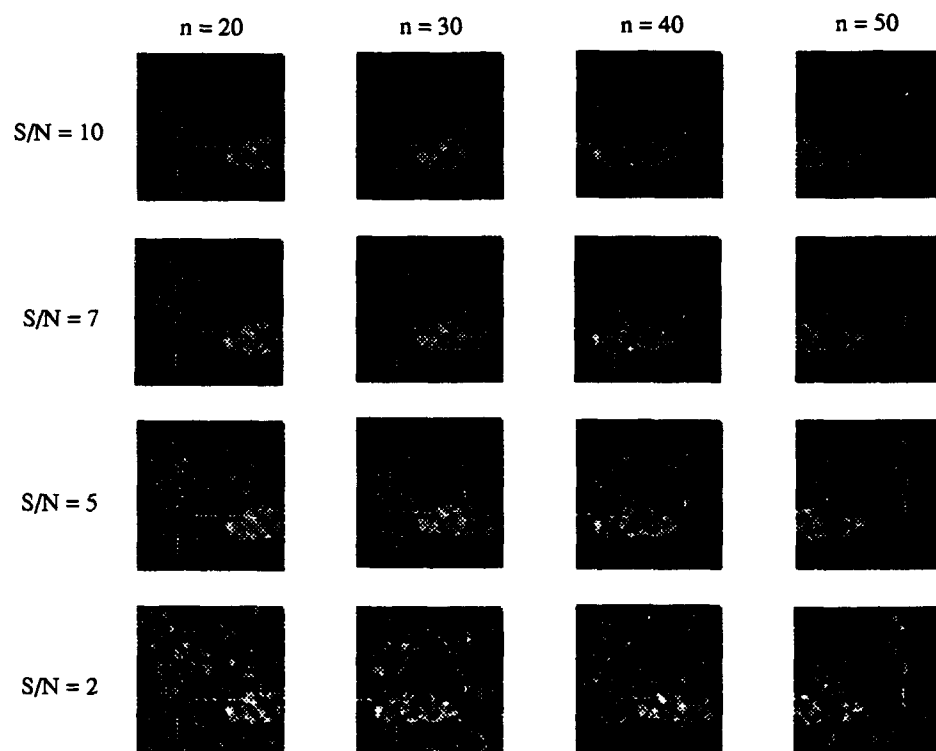


Figure 70. Four image sequences generated by adding equally correlated gaussian noise to the tank image sequence previously shown in Figure 58. The signal-to-noise ratio of each sequence is shown along the left hand side of the figure.

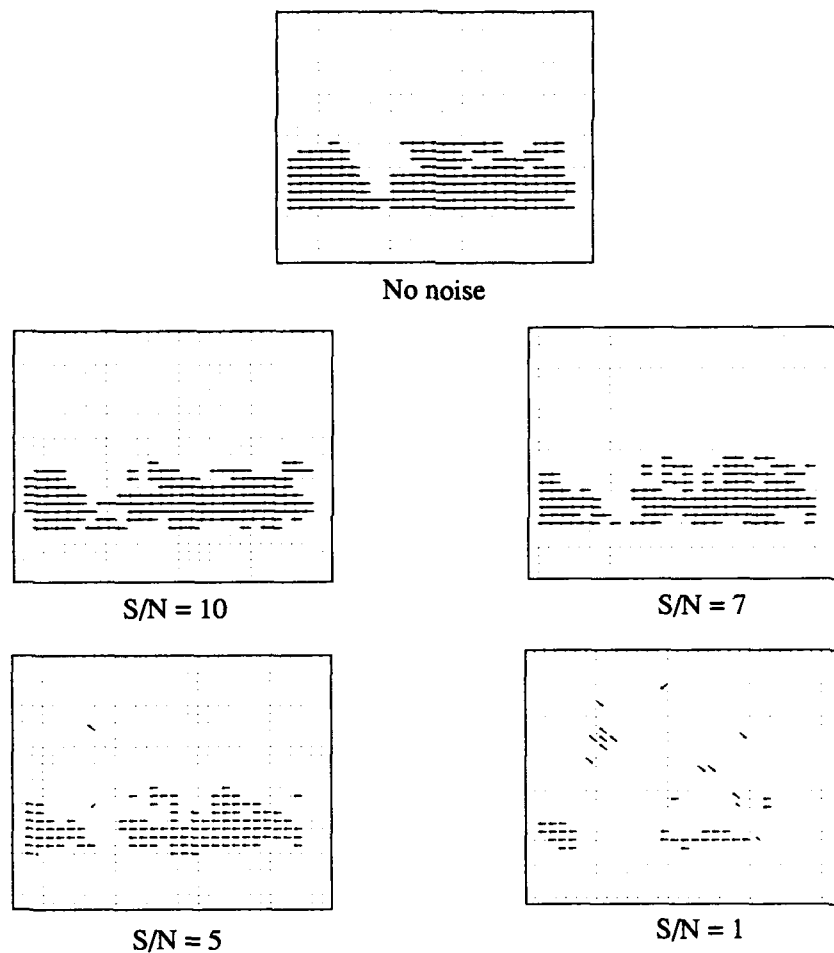


Figure 71. The upper flow map was produced by applying the flow computation and restoration algorithm to the original "noiseless" tank image sequence in Figure 58. The remaining flow maps correspond to various signal-to-noise ratios as indicated under each map.

four sets of coefficients that could be summed to capture the signal frequency content contained in one of four diagonally opposing regions in frequency space. The largest coefficients of the decomposition process at each location in the image sequence were then combined to compute the optical flow of the signal.

Several examples were provided which demonstrated the ability of the multiresolution wavelet vector motion analysis to correctly compute the optical flow of an image sequence and thereby discriminate between multiple objects of different sizes moving with different speeds and directions. Furthermore, like all optical flow algorithms, it was shown that the performance of the wavelet-based flow estimation algorithm was degraded by the presence of physical and system noise phenomena. The final section of the chapter therefore presented a unique flow restoration methodology that incorporated a modified version of S. Grossberg's gated dipole filter in a cooperative-competitive flow restoration methodology. Several examples were provided which demonstrated the ability of the flow restoration algorithm to find and correct localized flow inconsistencies caused by spatially correlated gaussian noise, occluding boundaries, and rapid fluctuations in reflected surface intensities.

The multiresolution wavelet vector motion analysis technique developed in this chapter offers several distinct advantages over other spatio-temporal frequency motion analysis approaches. First, and perhaps most important, the wavelet motion analysis provides a rigorous mathematical framework for the construction of a multiresolution, motion-oriented filter bank. Other spatio-temporal frequency approaches, most notably those developed by Heeger (24) and Watson (58), employ *ad hoc*, fixed window, Fourier filtering strategies that do not provide a formal mechanism to control key properties such as filter bandwidth, inter-band filter overlap, and space-time/frequency localization. Second, the 3D analyzing filters used in the motion analysis algorithm are constructed from the non-homogeneous multiresolution wavelet analysis developed in Chapter IV. Thus, their spatial and temporal frequency characteristics can be easily and independently varied to match the size and velocity constraints of the design scenario. And third, the pyramidal, sub-band coding scheme used to generate the motion-oriented filter bank provides a fast method for analyzing motion at multiple spatial scales.

Although the sub-band coding scheme used here is "fast" compared to the Fourier frequency filtering techniques used in other spatio-temporal frequency motion analysis approaches, it is still much too slow to implement in real-time on a sequential digital computer. For example, coding inefficiencies notwithstanding, the average processing time required to decompose a $128 \times 128 \times 128$ image sequence

across all spatial resolution levels, generate a flow map for each level and post-process the flow maps to remove flow inconsistencies was approximately 35 minutes on a single user SUN SPARCstation 2. Thus, the next chapter investigates the feasibility of reducing the computation time of the wavelet multiresolution analysis using two very different parallel architectures - a digital Hypercube, and an optical correlator.

VI. Increasing the Speed of the Motion-Oriented Multiresolution Wavelet Decomposition Algorithm through Digital and Optical Parallelization

6.1 Introduction

The bulk of the processing time required to run the wavelet vector motion analysis algorithm is taken up by the motion-oriented 3D wavelet decomposition process. Indeed, a *C* version of the $\mathcal{O}(N^3)$ serial decomposition algorithm requires approximately 30 minutes of wall clock time to fully decompose a $128 \times 128 \times 128$ image sequence on a dedicated SUN SPARCstation 2. Thus, the serial motion-oriented decomposition algorithm would be difficult to implement in real-time on existing single microprocessor platforms. The purpose of this chapter, therefore, is to investigate the potential for increasing the computational speed of the decomposition algorithm using digital and optical parallel architectures.

The first section of this chapter presents two parallel digital versions of the spatio-temporal decomposition algorithm developed and implemented by AFTT students on a SUN SPARCstation distributed network, an Intel i/PSC2 8-node Hypercube, and an i/PSC/860 64-node Hypercube. Experimental results demonstrate an approximately linear increase in decomposition speed with the number of Hypercube nodes. The second section of the chapter presents a parallel 2D optical wavelet architecture published as part of this research effort in *Optical Engineering*, September, 1992. The purpose of this phase of the research was to determine the feasibility of implementing the 2D spatial decomposition stage of the 3D algorithm using optical technology. Experimental results verify the feasibility of the concept; however, several adjustments must be made in the proposed architecture to make it applicable to a general class of wavelet filters.

6.2 Digital Parallelization

This section summarizes the results of an investigation conducted by members of the AFTT pattern recognition research group into the computational speed-up achieved by parallelizing the discrete motion-oriented wavelet decomposition process (53). Although the wavelet decomposition process is inherently parallelizable at many different scales, the original decomposition algorithm was implemented in *C* on a single microprocessor machine. This serial algorithm was divided along functional lines into modules that performed specific program tasks (e.g., load an image, perform

spatial decomposition, perform temporal decomposition, etc.). Captain Laura Suzuki and Lieutenant Rob Reid parallelized the serial program along these same functional lines, but at different levels of "granularity" designed to match the parallel processing capabilities of three different parallel systems: 1) a distributed SUN SPARCstation 2 network (coarse-grain), 2) an 8-node Intel iPSC/2 Hypercube, and 3) a 64-node Intel iPSC/860 Hypercube (fine-grain). Before discussing the parallel algorithms, a brief review is provided of the major functions of the serial decomposition algorithm as previously described in Chapter IV.

6.2.1 Serial Motion-Oriented Wavelet Decomposition Algorithm. A visualization of the key functions in the serial motion-oriented wavelet decomposition algorithm is shown in Figure 72. The input to the sequential algorithm is an $N \times N \times N$ (rows, column, frames) 3D image sequence representing the projection coefficients of the "zeroth" spatio-temporal approximation level. The first stage of the algorithm spatially decomposes each frame of the image sequence into the next lower spatial approximation and detail signals by convolving the 1D spatial scaling function and wavelet filters with the frame's rows and columns and keeping every other sample in space. This process is then recursively applied to each successively smaller spatial approximation signal until the number of samples in the spatial dimension are less than the number of coefficients in the spatial filter.

In the next stage of the algorithm, the spatial detail signals formed by each spatial decomposition are decomposed in time by convolving the 1D temporal wavelet and scaling function filters across all frames at each spatial location and keeping every other sample in time. Again, this process is recursively applied to each temporal approximation signal until the number of time samples in the final downsampled approximation signal is less than the number of coefficients in the temporal wavelet filter.

As derived in Chapter IV, the computational complexity of the discrete motion-oriented multiresolution wavelet decomposition algorithm is of $\mathcal{O}(N^3)$. This high degree of complexity makes it difficult to apply the algorithm to real-time target discrimination problems using off-the-shelf, single microprocessor platforms. Thus, the next section describes how the sequential algorithm described above was parallelized to run on a distributed SUN SPARCstation 2 network, and on two 8 and 64-node Hypercubes.

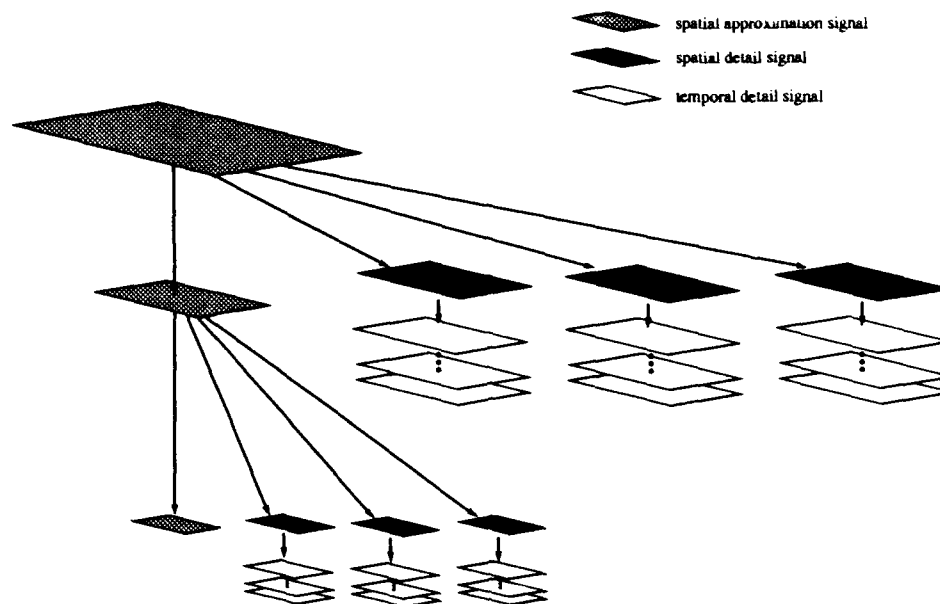


Figure 72. A visualization of the motion-oriented multiresolution wavelet decomposition process.

6.2.2 *Parallel Algorithms for Distributed SUN SPARCstation 2 Network and Intel iPSC/2 and iPSC/860 Hypercubes.* Viewed from a parallel programming standpoint, a major advantage of the motion-oriented wavelet decomposition algorithm is its scalability (53). That is, each of the coefficients produced by one spatial decomposition are computed independently, so that each computation can be assigned to a single parallel node. Thus, given an unlimited number of parallel processing nodes, one can theoretically reduce the order of the spatial decomposition from $\mathcal{O}(N^2)$ to $\mathcal{O}(\log_2 N)$, where $\log_2 N$ is the bound on the number of spatial decompositions. Unfortunately, the parallel processing systems available to AFIT have a limited number of nodes; thus, the parallel implementations described next were much more “coarse-grain” than the $\mathcal{O}(\log_2 N)$ scenario.

The first parallel design was implemented on a distributed SUN SPARCstation 2 network. This design, like the Hypercube design discussed next, broke out the program tasks along functional lines. A task graph showing the major functions of the serial 3D decomposition program *packet3D* is shown in Figure 73a). The lightly shaded circles represent the critical path of the decomposition. They include the initial task of reading in the 3D input sequence (1), and the follow-on tasks of generating a spatial approximation signal at each lower spatial decomposition level (3). Task (2) represents the generation of the three spatial detail signals, and task (4) represents the complete temporal decomposition process.

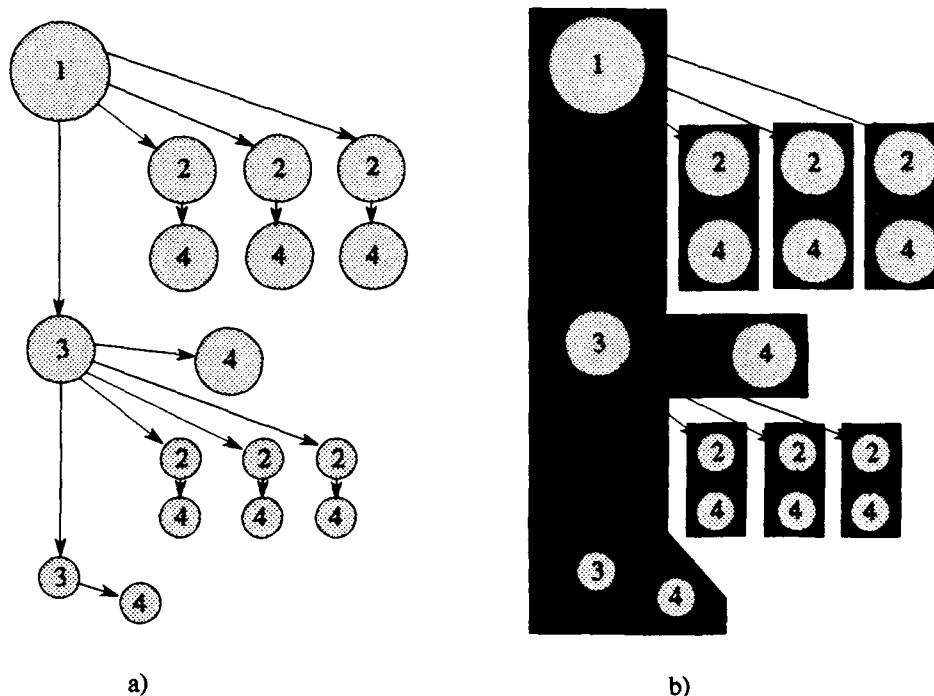


Figure 73. a) Major processing tasks associated with the motion-oriented wavelet decomposition algorithm. The lighter shaded circles represent the critical path of the algorithm. Task (1) reads in the input coefficients and generates the first spatial approximation level. Task (2) generates the three spatial detail signals. Task (3) generates successively smaller spatial approximation signals, and Task (4) generates a complete set of temporal detail signals. b) Task assignments for the distributed decomposition algorithm shown by blocked regions (53).

Figure 73b) displays the manner in which these tasks were assigned to the distributed SUN network. Here, the critical path is left running on a single machine, and additional tasks are allocated (or forked) to the remaining machines as data becomes available. One of the primary problems with this approach is that the number of parallel tasks allocated during the spatio-temporal decomposition process is fixed. Therefore, this coarse-grain parallel design is not scalable to the size of the input signal.

The second parallel design was implemented on an 8 and a 64-node Intel Hypercube. In this design, a single node designated as a system supervisor sends data and control instructions to each of the worker nodes. When the worker node has completed its task, it returns the results to the supervisor node and awaits further processing instructions. The algorithm begins by passing a different 2D frame of data to every available node. After receiving a frame of data, each node performs a complete 2D

spatial decomposition and returns the results to the supervisor. The supervisor node then recombines the data and initiates the temporal decomposition process. The temporal decomposition is accomplished by bundling together groups of "time-strings" and passing them to available nodes for processing. A time-string consists of the temporal values "found on a line drawn through all the frames at a given (x, y) point in some level of the spatial decomposition (53)." Since the degree of parallelism in the spatial and temporal decomposition processes is dependent on the number of frames and time-strings in the image sequence, this program is considerably more scalable than the distributed SUN algorithm. A drawback to this architecture, however, is the memory restrictions of the individual processing nodes limit the dimensions of the input image sequence to $48 \times 48 \times 48$. The next section describes the reduction in processing time achieved with both algorithms.

6.2.3 Tests and Results. Three experiments were conducted to determine the average speed-up of the distributed SUN and Hypercube algorithms over the serial version of the motion-oriented multiresolution wavelet decomposition algorithm. For the purposes of these tests, the speed-up, S is defined as (53)

$$S = \frac{\tau_{base}}{\tau_{run}} \quad (157)$$

where τ_{base} is the baseline run time of the serial algorithm, and τ_{run} is the run time of the parallel algorithm.

The first test was performed to determine the average speed-up of the distributed SUN algorithm. The test was performed on a $128 \times 128 \times 128$ image sequence using a 12 tap Daubechies' wavelet in space and time. The baseline time was determined by timing the serial version of the algorithm from start to finish. The average baseline time over five runs was 31.6 minutes. Since the decomposition algorithm was distributed across several machines, the run time was determined by measuring the time required to execute the critical path on the main machine, as well as the execution times of the side processes on the remaining three machines. Averaging the results over five runs, the overall speed up of the distributed network was

$$\begin{aligned} S_{overall} &= \min(S_{critical}, S_{side}) \\ &= \min\left(\frac{31.6}{10.2}, \frac{31.6}{16.6}\right) \\ &= 1.904 \end{aligned} \quad (158)$$

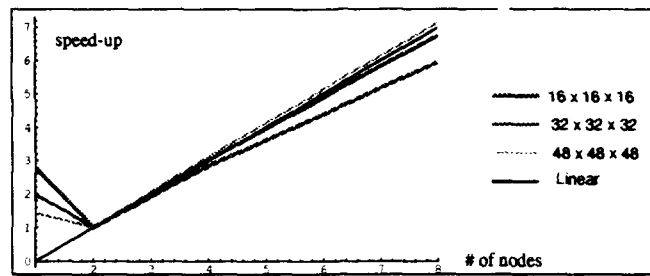


Figure 74. Speed-up vs. number of nodes for the iPSC/2 implementation using a Daubechies 8 tap filter in space and time (53).

which is less than the linear value of 4. This slower-than-linear behavior is caused by I/O contention created when the machines running the side processes are forced to wait to obtain input data from the same file server. Although not attempted here, it should be possible to alleviate this contention by passing the information directly to the side path machines. This would in turn reduce the run times of the side paths and increase the speed-up of the distributed SUN system.

The second experiment was performed using an 8-node Intel iPSC/2 Hypercube. Four test image sequences were used in the experiment. The dimensions of the test sequences were $8 \times 8 \times 8$, $16 \times 16 \times 16$, $32 \times 32 \times 32$ and $48 \times 48 \times 48$ (recall that memory limitations prevented the use of larger image sequences). The experiments were conducted with both a Daubechies' 4 tap and a Daubechies' 8 tap spatio-temporal filter. It was determined that the serial version of the algorithm was different enough from the parallel version that it made a poor baseline for the tests. Thus, a 2-node version of the Hypercube algorithm was chosen as the baseline test case. Since the memory allocation scheme used in these tests was inefficient, the run times were obtained by measuring "computational" times only (i.e., memory allocation and I/O times were ignored). Figure 74 shows the speed-up for 2, 4 and 8 nodes using the Daubechies 8 tap filter. Note that the speed-up is nearly linear for each of the test image sequences.

The third and final experiment was conducted with an iPSC/860 64-node Hypercube. Although the parallel algorithm was developed and debugged at AFTT, the tests were conducted on an iPSC/860 located in Beaverton Oregon. Additionally, with the exception of the number of nodes used in the experiments, the tests performed here were identical to those previously performed with the iPSC/2. The results of the iPSC/860 tests for an 8 tap spatio-temporal Daubechies filter are shown in Figure

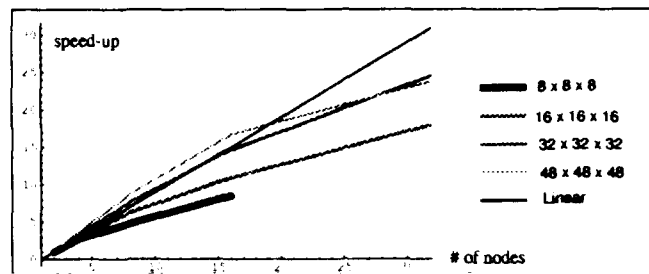


Figure 75. Speed-up vs. number of nodes for the iPSC/860 implementation using a Daubechies 8 tap filter in space and time (53).

75. Notice that the speed-up past the 8 – 10 node point is generally slower-than-linear. This behavior is caused by the fact that the communications time, which is included in the run time calculation and which is approximately constant for all configurations, is considerably larger than the nodal processing time. Thus, as more nodes are added to the system configuration, the reduction in processing time is small compared to the constant communications time and the speed-up no longer appears linear. However, the speed-up is still quite significant, particularly for the larger test sequences.

The results of the above experiments suggest that the time required to perform a 3D motion-oriented decomposition can be significantly reduced by parallelizing the algorithm and running it on a parallel digital platform such as the Intel Hypercube. Furthermore, it appears quite possible that one can improve the speed-up beyond the results achieved here by, for example, more efficiently allocating memory, reducing I/O operations, and lowering the communications overhead of the algorithm. In the next section, an optical alternative for increasing the speed-up is explored in which the spatial decomposition stages of the algorithm are implemented in frequency space using Fourier transforming lenses and thermoplastic holography.

6.3 Optical Parallelization

Test results from the previous section showed that the speed-up achieved by parallelizing the motion-oriented multiresolution wavelet algorithm on a digital Hypercube increases approximately linearly with the number of parallel nodes used in the computation. Theoretically then, one can achieve real time processing speeds by simply adding enough nodes to the parallel implementation. However, the speed-up curves obtained for the modestly sized $48 \times 48 \times 48$ image sequence show

that even if one ignores memory allocation and I/O times, thousands of nodes would be required to reach a minimum video system computational throughput of thirty frames per second. These types of frame rates would clearly be difficult to achieve with existing parallel digital technology. Thus, the purpose of this section is to present the results of an investigation into the feasibility of using a parallel optical architecture for performing the 2D spatial decompositions required by the first stage of the spatio-temporal decomposition algorithm.

6.3.1 Optical Wavelet Theory. Chapter II described how a continuous 1D wavelet transform can be implemented as a Fourier filtering operation. This property makes the 2D wavelet transform ideally suited for optical implementation. To demonstrate this property, let the two dimensional, spatial wavelet transform, $[W_\psi i](a, b, c, d)$, of an image, $i(x, y)$, be given by

$$[W_\psi i](a, b, c, d) = \int \int_{-\infty}^{+\infty} i(x, y) \frac{1}{\sqrt{ab}} \psi \left(\frac{x-c}{a}, \frac{y-d}{b} \right) dx dy \quad (159)$$

where a, b are dilation parameters, and c, d are translation parameters. Letting $\psi_{ab}(x, y) = \frac{1}{\sqrt{ab}} \psi \left(\frac{x}{a}, \frac{y}{b} \right)$, Equation 159 can be rewritten as

$$[W_\psi i](a, b, c, d) = \int \int_{-\infty}^{+\infty} i(x, y) \psi_{ab}(x - c, y - d) dx dy \quad (160)$$

Equation 160 shows the wavelet transform can be expressed as a correlation process in which the image is correlated with a dilated version, ψ_{ab} , of the wavelet ψ . Of course, correlations can be easily implemented as filtering operations in the spatial frequency domain. Thus, if $I(f_x, f_y) = \mathcal{F}\{i(x, y)\}$, $\Psi(f_x, f_y) = \mathcal{F}\{\psi(x, y)\}$, and $\Psi_{ab}(f_x, f_y) = \mathcal{F}\{\psi_{ab}(x, y)\}$, then Equation 5 becomes

$$[W_\psi i](a, b, c, d) = \left[\mathcal{F}^{-1} \{ I(f_x, f_y) \cdot \sqrt{ab} \Psi(-af_x, -bf_y) \} \right] (c, d) \quad (161)$$

where $\sqrt{ab} \Psi(af_x, bf_y) = \Psi_{ab}(f_x, f_y)$. For a given pair of dilation parameters a, b , Equation 161 can be implemented optically by first recording the image's Fourier transform on a hologram placed in the frequency plane of a Vander Lugt correlator (21). The hologram is then illuminated with the Fourier transform of an appropriately dilated Haar wavelet, and the resulting correlation term is Fourier transformed to produce the wavelet transform. The results described in this section were produced

using a similar approach in which the image and the optical wavelet were generated with a binary SEMETEX 128×128 MOSLM. (8).

One drawback to using a binary device to generate an optical wavelet is that it cannot produce the continuous scaling factor traditionally associated with a wavelet. For example, in one dimension the wavelet kernel is typically given by $\frac{1}{\sqrt{a}}\psi\left(\frac{x-b}{a}\right)$ where the scaling term $\frac{1}{\sqrt{a}}$ acts as an energy normalization factor (14). Because the wavelet and the image are both generated with the MOSLM, it is not possible to include an energy normalization factor in the optical wavelet transform. However, the absence of the normalization factor does not affect the existence of the transform's inversion integral (8); therefore, the optical implementation presented here is referred to as a "wavelet" transform, where the transform filtering operation in terms of the unnormalized wavelet kernel, $\psi\left(\frac{x-c}{a}, \frac{y-d}{b}\right)$, is given by

$$[W_h i](a, b, c, d) = [\mathcal{F}^{-1}\{I(f_x, f_y) \cdot ab\Psi(-af_x, -bf_y)\}](c, d) \quad (162)$$

A second drawback to generating an optical wavelet with a binary device is that one is restricted to wavelets with binary amplitude distributions. One such mother wavelet commonly used in early wavelet applications is the Haar wavelet (11). In two dimensions, a separable Haar mother wavelet is given by

$$h(x, y) = \begin{cases} 1 & \text{if } 0 \leq x < .5 \text{ and } 0 \leq y < .5 \text{ or } .5 \leq x < 1 \text{ and } .5 \leq y < 1 \\ -1 & \text{if } .5 \leq x < 1 \text{ and } 0 \leq y < .5 \text{ or } 0 \leq x < .5 \text{ and } .5 \leq y < 1 \\ 0 & \text{otherwise} \end{cases} \quad (163)$$

A two dimensional Haar mother wavelet and its Fourier transform are shown in Figure 76. Although the discontinuities along the borders between the zero, positive, and negative states of the Haar wavelet make it undesirable as a kernel for transforming highly continuous images, the uniform regions of intensity between these well defined discontinuities make it suitable for implementation with binary electro-optic devices. In particular, this section examines the feasibility of implementing a Haar wavelet transform using a Vander Lugt optical correlator in which a family of Haar wavelets is generated with a SEMETEX 128×128 Magneto-Optic Spatial Light Modulator (MOSLM).

Figure 77 contains a digital simulation of the optical Haar wavelet transform where the binarized, 128×128 image used in the experiment (Figure 77a) is correlated with four differently dilated Haar

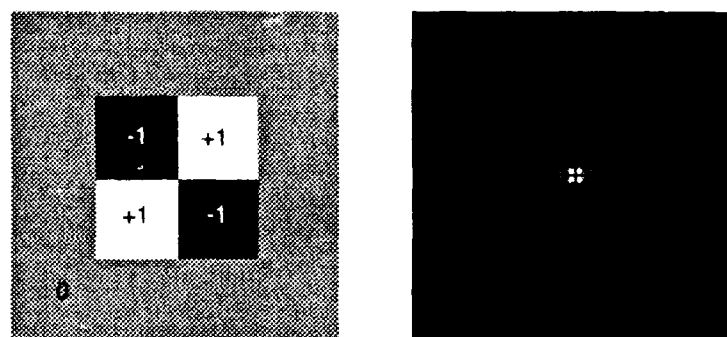


Figure 76. Two dimensional Haar mother wavelet (left) and its Fourier transform (right).

wavelets. In order to accurately simulate the holographic storage and recall process, each of the wavelet space projections was obtained by first multiplying the Fourier transform of an unnormalized Haar wavelet with the conjugate of the Fourier transform of the image, and then Fourier transforming the result. The optical designs used to accomplish this process are described in the following section.

6.3.2 System Design. Two system designs were used to optically implement Equation 162. Both designs used different methods for generating a family of Haar wavelets on the MOSLM. The first method employed the ternary phase-amplitude state capability of the MOSLM. Here, the $+1, 0, -1$ pixels of the Haar wavelet were generated by operating the MOSLM in a phase-only mode. The surrounding 0 state pixels were generated by accessing the MOSLM's neutral, or demagnetized state. In this state, plane wave light passing through the demagnetized pixels was diffracted into higher order spatial frequency components which were eliminated by low-pass spatial filtering techniques (35). Multiple dilations were achieved by electronically varying the number of $+1, -1$ and 0 state pixels.

The second method used a variable square aperture to control the dilation of the wavelet. Here, a full 128×128 Haar wavelet was written on the MOSLM and a variable square aperture positioned at the center of the wavelet controlled the wavelet's dilation. The 0 state was obtained by simply blocking the light surrounding the $+1, -1$ region inside the square aperture. This dilation technique was investigated after experiments revealed the ternary mode MOSLM was unable to produce a true zero state.



a) 128 x 128 image



b) 4 x 4 wavelet



c) 8 x 8 wavelet



d) 16 x 16 wavelet

Figure 77. Theoretical results from a digital simulation of the optical continuous Haar wavelet transform: a) 128×128 binarized input image. b) - d) Correlation of the input image with four different wavelet dilations.

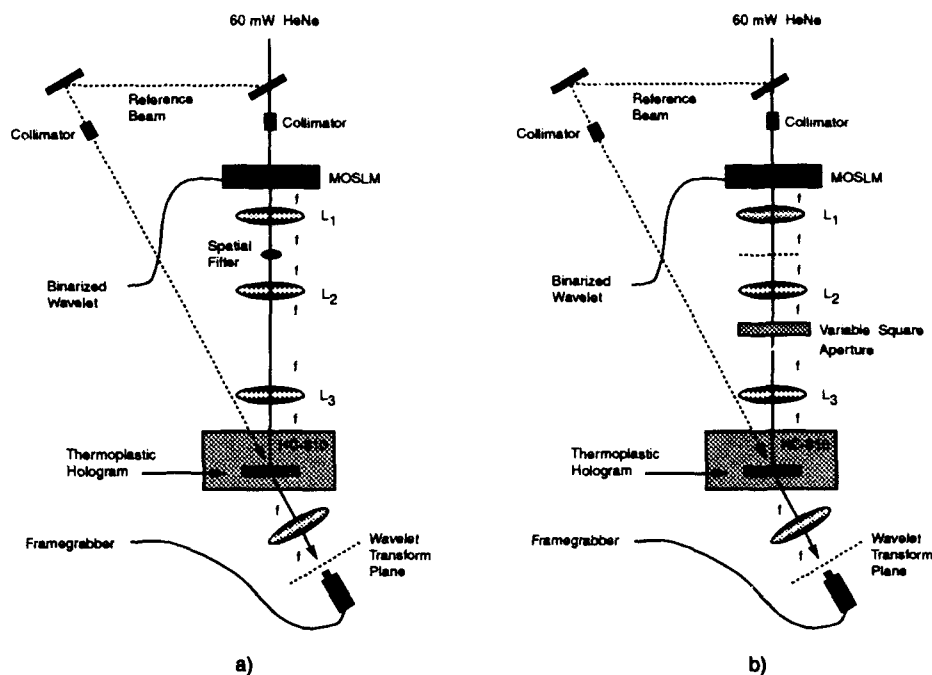


Figure 78. Vander Lugt optical correlation designs: a) Wavelet dilations controlled electronically using ternary state MOSLM and spatial filter. b) Wavelet dilations controlled by variable square aperture.

Figure 78 shows the Vander Lugt correlation schemes used to implement both wavelet transform methods. In each implementation, a 128×128 binarized input image was generated by the MOSLM and recorded on a thermo-plastic hologram using a Newport HC-310 thermal holographic camera. The reference-to-object beam power ratio generated by the 60mw HeNe laser was maintained at 10 : 1 and the reference beam angle was 31° off the optic axis. In the first design (see Figure 78a) two lenses, L_1 and L_2 , together with a low-pass spatial filter are used to block the energy diffracted into higher order spatial frequency components by the demagnetized pixels in the MOSLM. The second design, Figure 78b), uses lenses L_1 and L_2 to image the 128×128 Haar wavelet onto a variable square aperture located in the front focal plane of the $4f$ correlator. The correlation results for both designs were imaged onto a CCD array and captured using an AT&T Truevision Advanced Raster Graphics Adapter (TARGA) framegrabber.

6.3.3 Tests and Results. An example of a Haar wavelet generated using the first design method is shown in Figure 79. This figure was obtained by rotating the output polarizer of the MOSLM

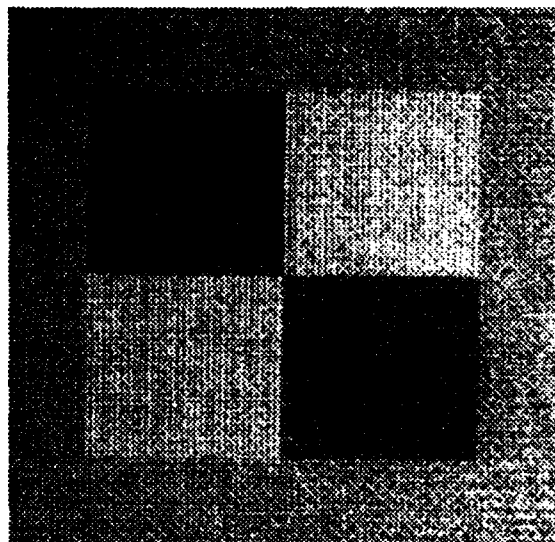


Figure 79. Optical Haar wavelet implemented with ternary state MOSLM. Center region containing 1, -1 pixels is 64×64 . Output polarizer is adjusted for maximum contrast.

to produce maximum contrast between its three operational states. Figure 80 shows a typical result obtained by correlating a 16×16 ternary phase-amplitude state wavelet with the binarized input image shown in Figure 77a). The poor quality of these results was traced to the inability of the spatial filtering process to remove enough diffracted energy from the zero state pixels for the MOSLM to adequately approximate the behavior of a Haar wavelet. Although several potential solutions to this problem were implemented, none yielded adequate results. Thus, even though the ability to electronically clock in multiple wavelet dilations using the ternary phase-amplitude mode of the MOSLM is highly desirable, this research shows the ternary design technique cannot be implemented satisfactorily with current MOSLM technology. The second design method yielded substantially better results.

Figure 81 shows a 32×32 Haar wavelet produced using the aperture stop wavelet design method. In this figure, the output polarizer is adjusted to produce maximum contrast between the two different wavelet states. During operation, the polarizer is adjusted so that the MOSLM operates in a phase-only mode to yield a uniform intensity across the +1 and -1 regions of the wavelet.

Figure 82 allows a comparison of the theoretical and experimental results obtained by correlating the input image with a 8×8 Haar wavelet. Although there are strong similarities in the shape and contrast of the edges predominantly detected by both the digital and optical wavelets, the digital wavelet



Figure 80. 16×16 spatially filtered wavelet correlated with binarized input image.

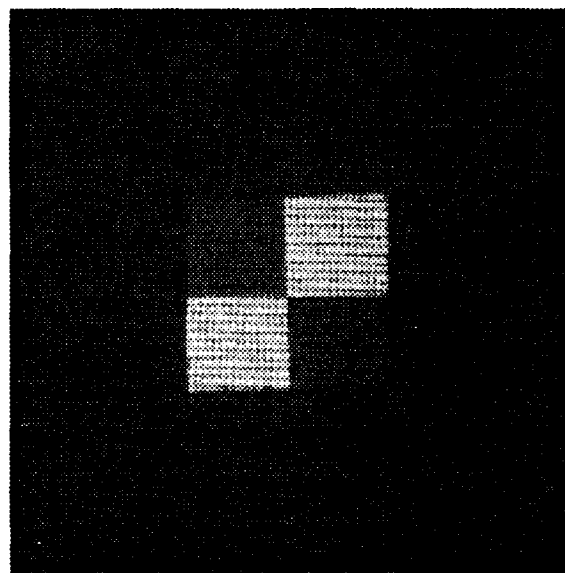


Figure 81. 32×32 Haar wavelet produced using the aperture stop wavelet design method.

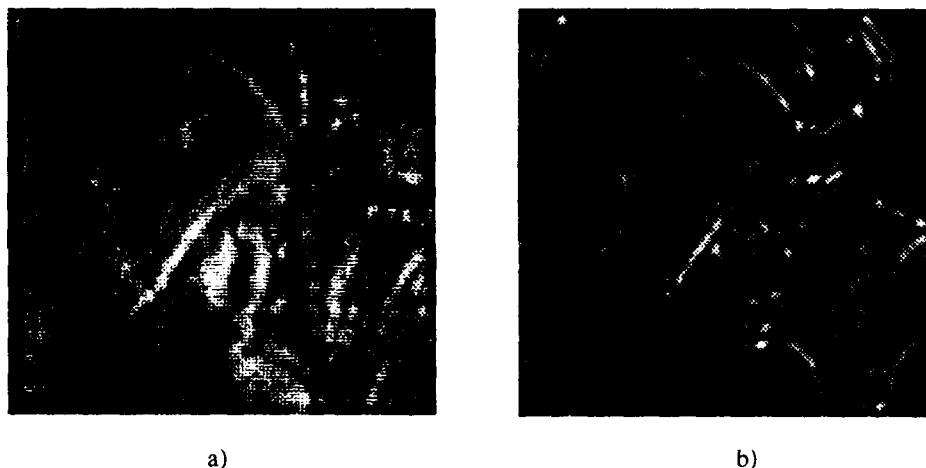


Figure 82. Comparison of digital a) and optical b) results obtained by correlating the binarized input image with an 8×8 Haar wavelet.

was able to resolve finer details in the scene. This difference is attributable to the existence of a DC component in the optical wavelet's frequency spectrum (recall that a true wavelet is a "zero mean" signal so that the wavelet filter has no DC frequency component). The DC component is caused by 1) SLM pixel drop-outs that prevent the number of $+1$ and -1 pixels in the wavelet from summing to zero, and 2) unavoidable stray light passed by approximately 25% of each of the SLM's pixels. Additionally, the frequency spectrum of the wavelet is altered by energy diffracted off the edges of the square aperture. These factors combine to allow low frequency components of the input image spectrum to survive the correlation process, thereby reducing the edge resolving capabilities of the optical wavelet transform.

6.4 Conclusion

Two different parallel methods were investigated for reducing the computational time of the order $O(N^3)$ serial, motion-oriented wavelet decomposition algorithm. The first, or digital, method employed three digital parallel architectures with varying degrees of "granularity." The first architecture was a coarse-grain parallelization of the serial algorithm using a distributed SUN SPARCstation 2 network. Although the overall processing time was reduced by a factor of approximately 2, the system and algorithm designs were not scalable. The second two architectures investigated were Intel iPSC/2 and iPSC/860 Hypercubes. The parallel algorithms associated with these platforms were less "coarse-

grained" than the distributed SUN algorithm in that they were scalable with the number of frames and "time-strings" in the input image sequence. The results of several tests conducted with differently sized image sequences and wavelet filters displayed a near linear speed-up in computational time over a serial baseline system. Furthermore, the results suggest that the speed-up can be further increased by improving administrative tasks such as memory allocation and I/O operations.

The second method presented was designed to investigate the feasibility of using an optical architecture to implement the spatial decomposition stages of the serial motion-oriented decomposition algorithm. Here, two different optical architectures were presented for implementing an optical Haar wavelet transform. Both methods used a Vander Lugt correlator to perform the wavelet transform and a SEMETEX 128×128 MOSLM to generate multiple dilations of the Haar mother wavelet. The most successful method employed a variable square aperture to control the dilation of a 128×128 Haar wavelet written to the MOSLM. This method was developed to compensate for the poor ternary phase-amplitude operation of the MOSLM. Although the results of the variable aperture method compare favorably to a digital simulation of the process, inherent limitations of the binary MOSLM (e.g., stray light, pixel drop-outs) prevent the creation of a true "zero mean", scaled wavelet. Additionally, the binary device does not allow for the construction of more general wavelets with continuous amplitude distributions.

VII. Conclusion and Contributions

7.1 Conclusion

Accurately detecting and discriminating multiple objects moving across a 2D sensor array in the presence of physical and system noise is an unsolved problem. Modern military target recognition systems generally employ a "static is basic" strategy in which single, static image frames from a time sequence of two dimensional IR imagery are individually analyzed for hot spots in the scene. However, biological studies show many animals rely on a "motion is basic" strategy in which objects are detected and identified by analyzing the temporal behavior (direction and speed) of more complex object attributes such as texture, orientation, edges and color. Furthermore, research into mammalian motion analysis systems reveal the existence of biological motion detectors that are 1) localized in space, 2) spatial frequency specific, and 3) sensitive to both direction and speed for spatial frequency contrasts greater than the subject's contrast sensitivity. The goal of this research, therefore, was to develop a computer vision-based motion analysis system that borrows on these biological concepts to discriminate between multiple objects moving in a noise corrupted scene. This goal was achieved through the development of a unique and powerful spatio-temporal, multiresolution wavelet motion analysis tool that computes the location, speed and direction of 2D brightness patterns moving within a sampled 3D image sequence.

Previous computer vision techniques designed for this same purpose have generally fallen into two categories: spatio-temporal differentiation, and, more recently, spatio-temporal integration techniques. Spatio-temporal differentiation techniques were made popular from the discovery that localized velocity information can be computed from the spatio-temporal gradient of a moving brightness pattern. However, these techniques require densely sampled imagery and are extremely sensitive to noise. They also require the use of *ad hoc* rules to compute motion at object boundaries and in regions of constant intensity. Spatio-temporal integration techniques, on the other hand, deduce local motion by integrating across many frames in an image sequence, generally in the form of a convolution or Fourier filtering operation. An important advantage of spatio-temporal integration is that flow discontinuities caused by object boundaries, occlusions and image noise are "averaged out" in the integration process. The disadvantage of previous spatio-temporal integration methods is that they employ heuristic mathematical techniques that provide little control over inter-dependent filtering characteristics such as filter overlap,

filter bandwidth, and space-time/frequency localization. Additionally, these techniques use rigid filter designs that cannot be easily modified to meet a particular problem scenario. The spatio-temporal integration technique presented here solves these and other problems by constructing a motion analysis framework on the rigorous mathematical foundation provided by a multiresolution wavelet analysis.

The development of the wavelet motion analysis tool occurred in a stepwise fashion, where each step of the research depended on the success of the previous step. Thus, although the overall contribution of this research effort is the development of mathematically rigorous yet flexible spatio-temporal frequency motion analysis tool, several smaller contributions were made throughout the design process. These contributions are described below in the order which they occurred in the design process.

7.2 Contributions

- **An $L_2(\mathbb{R}^3)$ Wavelet Multiresolution Analysis.** Although Y. Meyer developed the general theory for wavelet multiresolution analyses in $L_2(\mathbb{R}^n)$, previous instantiations of the wavelet multiresolution analysis dealt exclusively with one and two-dimensional signals. Thus, the first step in this research effort was to extend the mathematical details that governed the construction of wavelet orthonormal bases for $L_2(\mathbb{R})$ and $L_2(\mathbb{R}^2)$, to the space of finite energy spatio-temporal signals, $L_2(\mathbb{R}^3)$. As a result of this effort, it was shown that a separable wavelet orthonormal basis for $L_2(\mathbb{R}^3)$ consists of a set of seven dyadic wavelets evaluated over all possible integer shifts and dilations. Additionally, an "oct-tree" sub-band coding scheme for implementing a "Discrete Spatio-temporal Wavelet Transform" was developed which generates a bank of non-overlapping octave-band filters with uniform spatial and temporal frequency characteristics. This sub-band decomposition algorithm was then applied to a synthetic 3D image sequence to demonstrate its ability to extract vertical, horizontal or diagonal features from moving or stationary targets.
- **A Non-Homogeneous $L_2(\mathbb{R}^3)$ Wavelet Multiresolution Analysis.** The approximation spaces of the $L_2(\mathbb{R}^2)$ wavelet multiresolution analyses developed by previous researchers, as well as the above $L_2(\mathbb{R}^3)$ wavelet multiresolution analyses, are formed from the tensor product of identical 1D approximation spaces. This approach produces a scaling function filter with identical passband characteristics in each dimension of Fourier frequency space. This in turn limits the filter designer's ability to tailor the frequency characteristics of the wavelet filter to

match the frequency behavior of the input signal. Additionally, it reduces the computational efficiency of the 2D or 3D discrete wavelet transform algorithm by preventing the use of lower order, less computationally expensive filters along frequency coordinates with "looser" design constraints (e.g., a motion analysis problem that requires a higher degree of spatial frequency resolution than temporal frequency resolution). The next step of the research therefore was to increase the flexibility of the wavelet filter design process to allow for the construction of a separable spatio-temporal wavelet filter with non-uniform spatial and temporal frequency characteristics. This was accomplished through the creation of a unique "non-homogeneous" $L_2(\mathbb{R}^3)$ wavelet multiresolution analysis which generates a separable wavelet orthonormal basis for $L_2(\mathbb{R}^3)$ consisting of seven dyadic 3D wavelets constructed from non-identical 1D spatial and temporal wavelets. The resulting theory is quite flexible in that it allows one to construct an orthonormal wavelet bases for $L_2(\mathbb{R}^3)$ from any three $L_2(\mathbb{R})$ scaling functions provided each generates an $L_2(\mathbb{R})$ multiresolution analysis.

- **A Motion-Oriented Multiresolution Wavelet Analysis: Decoupling the Spatial and Temporal Decomposition Processes.** At each stage in the "conventional" non-homogeneous 3D wavelet decomposition algorithm, the spatial and temporal samples of the approximation and detail signals are both equally decimated to yield a bank of analysis filters whose spatial and temporal bandwidths both decrease by a factor of two from one stage of the decomposition to the next. Thus, at any level in the decomposition process, one is required to analyze the signal at equal scales in space and time. It is therefore not possible with the conventional structure to generate a wavelet filter that captures the energy of moving objects with dissimilar spatial and temporal frequency characteristics such as large, fast objects (i.e., objects with high temporal frequency and low spatial frequency content), or small, slow objects (low temporal frequency and high spatial frequency content). The purpose of this phase of the research, then, was to develop a 3D multiresolution wavelet decomposition technique that provides the ability to independently zoom-in and zoom-out on spatial and temporal details in a 3D image sequence. This was accomplished by "decoupling" the spatial and temporal decomposition processes to produce a rich set of independent *spatio-temporally* oriented frequency channels for analyzing the size and speed characteristics of moving objects. The motion-oriented wavelet decomposition algorithm was applied to a battlefield IR image sequence which demonstrated its ability to

locate objects at different spatial scales in the presence of extraneous motion related phenomena such as camera jitter, background noise and sensor noise.

- **A *Vector Wavelet Motion Sensor*.** The motion-oriented multiresolution wavelet analysis described above was designed to detect objects of different sizes moving with different speeds across a two-dimensional image plane. The symmetric 3D filters produced by the decomposition process thus act as a scalar motion sensing detectors in that they respond to the magnitude of an object's velocity vector (i.e., its speed), rather than to the vector quantity of speed *and* direction. The purpose of this stage of the research, therefore, was to expand the properties of the motion-oriented wavelet analysis to provide a multiresolution motion analysis tool that discriminates multiple moving objects in a three-dimensional image sequence based on their location, size, speed and direction of motion. This effort was implemented by dividing a symmetric wavelet filter into four diagonally opposing frequency pairs whose response more accurately determines the speed and direction of a moving object. The method employs the unique concept of an "extended real signal", which, when decomposed under a motion-oriented multiresolution wavelet analysis, yields four sets of wavelet coefficients that can be summed to extract the portion of the signal's frequency spectrum that lies in a given diagonally opposing region in 3D frequency space. The spatio-temporal center frequencies of multiple diagonally opposing wavelet pairs are then used to compute the optical flow of spatio-temporal image sequence.
- **A *Cooperative-Competitive Optical Flow Restoration Mechanism*.** Like all optical flow algorithms, the performance of the wavelet-based flow estimation algorithm developed under this research effort is degraded by the presence of physical and system noise phenomena. Therefore, a unique flow restoration methodology was developed in the next phase of the research that incorporates a modified version of S. Grossberg's gated dipole filter in a cooperative-competitive flow restoration methodology that reinforces consistent flow behavior and removes flow inconsistencies. Several examples were provided which demonstrated the ability of the flow restoration algorithm to find and correct localized flow inconsistencies caused by spatially correlated gaussian noise, occluding boundaries, and rapid fluctuations in reflected surface intensities.
- **Digital and Optical Parallelization Techniques for Increasing the Speed of the Motion-Oriented Wavelet Decomposition Algorithm.** The bulk of the processing time required to run

the wavelet vector motion analysis algorithm is taken up by the motion-oriented 3D wavelet decomposition process. For example, coding inefficiencies notwithstanding, approximately 30 minutes of wall clock time was required to fully decompose a $128 \times 128 \times 128$ image sequence on a dedicated SUN SPARCstation 2. Thus, the serial motion-oriented decomposition algorithm would be difficult to implement in real-time on existing single microprocessor platforms. The purpose of this stage of the research, therefore, was to investigate the potential for increasing the computational speed of the decomposition algorithm using digital and optical parallel architectures. The contributions made in these two areas are discussed below.

1. *Digital Parallelization of the Discrete Motion-Oriented Multiresolution Wavelet Algorithm.* A serial C version of the 3D motion-oriented wavelet decomposition algorithm was parallelized by members of the AFTT pattern recognition group to investigate the speed-up potential of three digital parallel architectures with varying degrees of "granularity." The first architecture was a course-grain parallelization of the serial algorithm using a distributed SUN SPARCstation 2 network. Although the overall processing time was reduced by a factor of approximately 2, the system and algorithm designs were not scalable. The second two architectures investigated were Intel iPSC/2 and iPSC/860 Hypercubes. The parallel algorithms associated with these platforms were less "coarse-grained" than the distributed SUN algorithm in that they were scalable with the number of frames and "time-strings" in the input image sequence. The results of several tests conducted with differently sized image sequences and wavelet filters displayed a near linear speed-up in computational time over a baseline serial platform. Furthermore, the results suggest that the speed-up can be further increased by improving administrative tasks such as memory allocation and I/O operations.
2. *Optical Parallelization of a 2D Spatial Wavelet Decomposition.* The purpose of this research was to determine the feasibility of using a parallel optical architecture for performing the 2D spatial decompositions required by the first stage of the spatio-temporal decomposition algorithm. Here, two different optical architectures were investigated for implementing an optical Haar wavelet transform. Both methods used a Vander Lugt correlator to perform the wavelet transform and a SEMETEX 128×128 MOSLM to generate multiple dilations of the Haar mother wavelet. The most successful method employed a

variable square aperture to control the dilation of a 128×128 Haar wavelet. The results of this phase of the doctoral research are published in the September 1992 issue of *Optical Engineering*.

Bibliography

1. Adelson, E.H. and J.R. Bergen. "Motion channels based on spatiotemporal energy," *Invest. Ophthalmol. Vis. Sci. Suppl.*, 25 (1984).
2. Adelson, E.H. and J.A. Moshovon. "Phenomenal coherence of moving visual patterns," *Nature* 300, 523-525 (1982).
3. Adiv, Gilad. "Determining three-dimensional motion and structure from optical flow generated by several moving objects," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 7:384-401 (July 1985).
4. Aggarwal, J.K. and N. Nandhakumar. "On the computation of motion from sequences of images-A review," *Proc. of the IEEE*, 76:917-935 (August 1988).
5. Ballard, Dana H. and Christopher M. Brown. *Computer Vision*. Englewood Cliffs, New Jersey: Prentice-Hall Inc., 1982.
6. Barron et. al., J.L. "Performance of optical flow techniques," *IEEE Proc. Comp. Vis. Patt. Rec.*, 236-241 (1992).
7. Bhanu, Bir and Terry L. Jones. "Image understanding research for automatic target recognition," *Proceedings: DARPA Image Understanding Workshop*, 249-254 (January 1992).
8. Burns, T.J. et. al. "Optical haar wavelet transform," *Optical Engineering*, 31:1852-1857 (September 1992).
9. Burns, T.J. et. al. "Spatio-temporal signal processing using a discrete 3D multiresolution wavelet analysis," *SPIE Proceedings Intelligent Robots & Computer Vision XI: Biological, Neural Net. and 3D Methods*, 1826 (November 1992).
10. Burt, Peter J. "Smart sensing within a pyramid vision machine," *Proceedings of the IEEE*, 76:1006-1015 (August 1988).
11. Chui, Charles K. *Wavelets: a tutorial in theory and application*. San Diego: Academic Press, 1992.
12. Coifman, R.R. and Y. Meyer. *Size properties of wavelet-packets*, in *Wavelets and their applications*, Ruskai et al.. San Diego: Academic Press, 1992.
13. Daubechies, Ingrid. "Orthonormal bases of compactly supported wavelets," *Commun. Pure Appl. Math.*, 41:909-996 (November 1988).
14. Daubechies, Ingrid. *Ten lectures on wavelets*. Philadelphia: Society for Industrial and Applied Mathematics, 1992.
15. Daugman, J.G. "Networks for image analysis: motion and texture," *Proceedings IJCNN* (June 1989).
16. Dudgeon, C. E. and Russel M. Mersereau. *Multidimensional digital signal processing*. Englewood Cliffs, New Jersey: Prentice-Hall Inc., 1984.
17. Fielding, K. et. al. "Spatio-temporal pattern recognition using hidden Markov models," *Proceedings of the SPIE*, 2032 (July 1993).
18. Gabor, D. "Theory of Communication," *Journal of the IEE*, 93:429-457 (1946).

19. Gafni, H. and Y. Y. Zeevi. "A model for processing of movement in the visual system," *Biological Cybernetics*, 32:165-173 (1979).
20. Gonzales, Rafael C. and Paul Wintz. *Digital image processing*. Reading, Massachusetts: Addison-Wesley, 1987.
21. Goodman, J. *Introduction to fourier optics*. New York, New York: McGraw-Hill Book Comp., 1968.
22. Grossberg, Stephen and Ennio Mingolla. "Neural Dynamics of Surface Perception: Boundary Webs, Illuminants, and Shape-From-Shading," *Computer Vision, Graphics, and Image Processing*, 37:116-165 (1987).
23. Hang, Hsueh-Ming *et. al.* "Motion estimation using frequency components," *SPIE Proceedings*, 1818:74-84 (November 1992).
24. Heeger, D. "Model for the extraction of image flow," *Journal of the Optical Society of America*, 4:1455-1471 (August 1987).
25. Heeger, D. "Optical flow using spatiotemporal filters," *International Journal of Computer Vision*, 279-302 (1988).
26. Hildreth, Ellen Catherine. *The measurement of visual motion*. Cambridge, Massachusetts: The MIT Press, 1983.
27. Horn, Berthold Klaus Paul. *Robot vision*. Cambridge, Massachusetts: The MIT Press, 1986.
28. Horn, K. P. and E.J. Weldon Jr. "Direct methods for recovering motion," *Intl. Journal of Computer Vision*, 2:51-73 (1988).
29. Huang, T.S. and A.N. Nevrtali. "Motion and structure from feature correspondences: a review," *To Appear in Proceedings IEEE* (1993).
30. Huang, T.S. and R.Y. Tsai. *Image sequence analysis: motion estimation in Image Sequence Analysis*, T.S. Huang, Ed.. New York: Springer-Verlag, 1981.
31. Huang, T.S. *et. al.* "Interframe coding with general two-dimensional motion compensation.," *Proceedings ICASSP*, 464-466 (1982).
32. Ishiguro, T. and K. Iinuma. "Television bandwidth compression transmission by motion-compensated interframe coding," *IEEE Communications Magazine*, 24-30 (November 1982).
33. Jacobson, Lowell and Harry Wechsler. "Derivation of optical flow using a spatiotemporal-frequency approach," *Comp. Vis., Graphics and Image Proc.*, 38:29-65 (1987).
34. Jones, Alain L.M. "Wavelet filter visualization using volume rendering and animation," *Air Force Institute of Technology Internal Report*, CSCE 682 (March 1993).
35. Kast *et al.*, Brian A. "Implementation of ternary phase amplitude filters using a magneto-optic spatial light modulator," *Applied Optics*, 28:1044-1046 (March 1989).
36. Lim, J.S. *Two-dimensional signal and image processing*. Englewood Cliffs, New Jersey: Prentice-Hall Inc., 1984.
37. Livingstone, Margaret S. "Art, illusion and the visual system," *Scientific American*, 78-85 (January 1988).

38. Longuet-Higgins, H. C. *et. al.* "The interpretation of a moving retinal image," *Proc. Royal Soc. London B.*, 208:385-397 (July 1980).
39. Mallat, Stephane G. "Multiresolution approximation and wavelets," *Trans Amer. Math Soc.*, 69-88 (September 1989).
40. Mallat, Stephane G. "A theory for multi-frequency signal decomposition," *IEEE Trans. on Pattern Anal. and Mach. Intell.*, 11:674-693 (July 1989).
41. McKee, S.P. "A local mechanism for differential velocity detection," *Vision Res.* 21, 491-500 (1981).
42. Meyer, Yves. "Principe d'incertitude, bases hilbertiennes et algebres d'operateurs," *Seminaire Bourbaki*, 662:209-223 (October 1986).
43. Mishkin, Mortimer *et. al.* "Object vision and spatial vision: two cortical pathways," *TINS*, 414-417 (October 1983).
44. Murray, David W. and Bernard F. Buxton. "Scene segmentation from visual motion using global optimization," *IEEE Trans. on Patt. Anal. and Mach. Int.*, 9:220-228 (March 1987).
45. Naylor, Arch W. and George R. Sell. *Linear operator theory in engineering and science*. New York: Springer-Verlag, 1982.
46. Oppenheim, A. V. and R.W. Schaffer. *Discrete-time signal processing*. Englewood Cliffs, New Jersey: Prentice-Hall Inc., 1989.
47. Poggio, Tomaso *et. al.* "Computerized vision and regularization theory," *Nature*, 317:314-319 (September 1991).
48. Rioul, Oliver and Martin Vetterli. "Wavelets and signal processing," *IEEE SP Magazine*, 14-38 (October 1991).
49. Robson, J.G. "Spatial and temporal contrast-sensitivity functions of the visual system," *J. Opt. Soc. Am.* 56, 1141-1142 (1966).
50. Schanke, Kenneth. "Forced correlate," *USAF Fighter Weapons Review*, 26-29 (Summer 1992).
51. Seibert, Michael and Allen M. Waxman. "Adaptive 3-D Object Recognition from Multiple Views," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(2):107-124 (1992).
52. Sekular, Robert *et. al.* *The perception of motion*, in *Visual perception. The neurophysiological foundations*, Spillman *et al.*. New York: Academic Press, 1990.
53. Suzuki, Laura R. C. and Rob Reid. "Parallelization of a 3 dimensional wavelet decomposition," *Air Force Institute of Technology Internal Report, CSCE 656* (June 1993).
54. Taylor, A. and D. Lay. *Introduction to Functional Analysis*. Englewood Cliffs, New Jersey: John Wiley and Sons, 1982.
55. Ullman, Shimon. *The interpretation of visual motion*. Cambridge, Massachusetts: The MIT Press, 1979.
56. Vaidyanathan, P. P. *Multirate systems and filter banks*. Englewood Cliffs, New Jersey: P T R Prentics Hall, 1993.

57. Watson, Andrew B. and Jr. A.J. Ahumada. "A look at motion in the frequency domain in SIC-GRAPH/SIGART Interdisciplinary Workshop, Motion: Representation and Perception," (April 4-6 1983).
58. Watson, Andrew B. and Jr. A.J. Ahumada. "Model of human visual-motion sensing," *J. Opt. Soc. Am.* 2, 322-341 (1985).
59. Weeks *et. al.*, Arthur R. "Computer-generated noise images for the evaluation of image processing algorithms," *Optical Engineering*, 982-992 (May 1993).
60. Zeimer, R. E. and W.H. Tranter. *Principles of Communications*. Boston: Houghton Mifflin Company, 1990.
61. Zeki, Semir. "The visual image in mind and brain," *Scientific American*, 69-76 (September 1992).

Vita

Captain Thomas J. Burns was born on 19 January, 1957 in Denver, Colorado. He graduated from Iver C. Ranum High School in Westminster, Colorado in 1975 and attended Ohio State University where he received a Bachelor of Arts Degree in Philosophy in 1980. Upon graduation, he was accepted into the Air Force Undergraduate Engineering Conversion Program where he received his commission from Officer Training School and returned to Ohio State University to receive his Bachelor of Science Degree in Electrical Engineering. His first assignment was to Headquarters, Air Force Logistics Command at Wright-Patterson AFB, OH, where he served as Staff Electronics Engineer until entering the Air Force Institute of Technology (AFIT) in May 1989. In December 1990, he was awarded the M.S.E.E. Degree with an emphasis in electro-optics and pattern recognition. After receiving the M.S.E.E. degree, he was immediately extended into the AFIT, in-residence PhD program to continue his research in the pattern recognition area. His doctoral research focuses on the discrimination of moving targets in time sequential imagery.

Permanent address: 2015 W. Choctaw Dr.
London, OH 43140

REPORT DOCUMENTATION PAGE			Form Approved OMB No. 0704-0188	
1. AGENCY USE ONLY (Leave blank) 2. REPORT DATE December 1993 3. REPORT TYPE AND DATES COVERED Doctoral Dissertation				
4. TITLE AND SUBTITLE A Non-Homogeneous, Spatio-Temporal Wavelet Multiresolution Analysis and its Application to the Analysis of Motion			5. FUNDING NUMBERS	
6. AUTHOR(S) Thomas Joseph Burns, Capt, USAF			7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Air Force Institute of Technology, WPAFB OH 45433-6583	
8. PERFORMING ORGANIZATION REPORT NUMBER AFIT/DS/ENG/93D-12			9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) WL/AARA (Greg Powers) 2010 5th St. WPAFB, OH 45433-7001	
10. SPONSORING/MONITORING AGENCY REPORT NUMBER			11. SUPPLEMENTARY NOTES	
12a. DISTRIBUTION AVAILABILITY STATEMENT Approved for Public Release; Distribution Unlimited.			12b. DISTRIBUTION CODE	
13. ABSTRACT (Maximum 200 words) This research presents a multiresolution wavelet analysis tool for analyzing motion in time sequential imagery. A theoretical framework is developed for constructing an $L_2(R^3)$ wavelet multiresolution analysis from three non-identical spatial and temporal $L_2(R)$ wavelet multiresolution analyses. This framework provides the flexibility to tailor the spatio-temporal frequency characteristics of the three dimensional wavelet filter to match the frequency behavior of the analyzed signal. An unconventional, discrete multiresolution wavelet decomposition algorithm is developed which yields a rich set of independent spatio-temporally oriented frequency channels for analyzing the size and speed characteristics of moving objects. Unlike conventional wavelet decomposition methods, this algorithm provides independent zoom-in and zoom-out capability in space and time. Symmetric 3D filters produced by the unconventional decomposition process are combined with the properties of the Hilbert transform to produce a bank of directionally selective wavelet filters. Multiple directionally selective wavelet filters are integrated to form a multiresolution vector wavelet motion sensor capable of unambiguously computing the optical flow of a 3D image sequence. A unique flow restoration methodology is presented which incorporates a modified version of Grossberg's gated dipole filter in a cooperative-competitive flow restoration methodology that reinforces consistent flow behavior and removes flow inconsistencies. Finally, several digital and optical parallel architectures are investigated for their ability to speed up the 3D wavelet decomposition process.				
14. SUBJECT TERMS Wavelets, Motion Analysis, Signal Processing, Harmonic Analysis			15. NUMBER OF PAGES 170	
16. PRICE CODE			17. SECURITY CLASSIFICATION OF REPORT Unclassified	
18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified			19. SECURITY CLASSIFICATION OF ABSTRACT Unclassified	
20. LIMITATION OF ABSTRACT UL				