WAVELET CHARACTERISTICS OF EARLY VISION

Geoffrey Brooks
WL/MNGA, Eglin AFB FL 32542

ABSTRACT

Recent advances in wavelet theory are affording great opportunities for signal processing applications. Natural neuronal networks exhibit wavelet behavior from which structural and functional paradigms could be exploited for machine-vision applications. Provided here is a summary of the ways vertebrate vision systems naturally exhibit wavelet characteristics.

INTRODUCTION

Wavelets can be generally defined as little waves that start and stop and originate from a single basic function [18]. The Wavelet Transform decomposes signals into their wavelet components as the Fourier Transform decomposes signals into their frequency components. Fourier time-frequency representations are presented here as an introduction to the more general wavelet representations.

Early vision can be defined as the processes that recover the properties of object surfaces from 2D intensity arrays [6]. The structure and function of natural vision systems exhibit wavelet characteristics in many ways. The focus here is on vertebrate vision information pathways that begin in the retina and terminate in cortical processing stages. Many of these concepts are also common in insect vision.

CONVENTIONAL TIME-FREQUENCY REPRESENTATIONS

Fourier developed a series of weighted sine and cosine terms to represent a periodic waveform $f(t)$, with period $T = 2\pi/\omega_0$, where ω_0 is the fundamental radian frequency. The Fourier Series is an infinite sum of components, weighted by coefficients a_n , at integer, n , multiples of ω_0 [15,20]:

$$f(t) = \sum a_n e^{j\omega_0 n t} \quad (1)$$

As T increases, ω_0 decreases along with the frequency spacing between Fourier Series components (or terms). As $T \rightarrow \infty$, the frequency distance between components becomes infinitesimal so that $\omega_0 \rightarrow \omega$, a continuous variable. The

solution for coefficients, when integrated over one sample period, becomes the Fourier Transform:

$$(Ff)(\omega, t) = \int dt f(t) e^{-j\omega t} = \langle f(t), e^{j\omega t} \rangle, \quad (2)$$

where $\langle \cdot \rangle$ denotes the inner product.

For real applications, functions of time are analyzed in finite intervals. To transform $f(t)$ into its frequency components, the signal is assumed periodic in the sampling interval. The duration of this interval becomes the fundamental period T , which also defines the frequency resolution $\Delta\omega = \omega_0$. For long sampling intervals (large T), frequency details are well resolved (small $\Delta\omega$), but the location in time of specific events within T is not known. Compensating for this by reducing T around a specific time t_0 results in large $\Delta\omega$, hence an uncertainty of specific frequency components. This tradeoff between time and frequency resolution is known as the *time-frequency uncertainty*.

An alternative to a specified duration of $f(t)$ is to window $f(t)$ with a Gaussian-like function, $g(t)$, centered at time t_0 . This windowing operator, T^{win} , serves to weight the signal closest to time t_0 more heavily than the signal farther away in time while maintaining a reasonably large T for adequate frequency resolution. In addition, the decay of T^{win} reduces the undesirable effects of *Gibb's phenomenon* at the edges [15]. The result is a Fourier transform that is a function of t_0 , called the *Windowed Fourier Transform*:

$$(T^{win}f)(\omega, t_0) = \int dt f(t) g(t-t_0) e^{-j\omega t} = \langle f(t), g(t-t_0) e^{j\omega t} \rangle \quad (3)$$

The *Discrete Windowed Fourier Transform* results in integer m sampled spectral components ($m\omega_0$) derived from integer n sampled time intervals (nt) of a signal. It is given as [7]

$$T^{win}_{m,n}f = \int dt f(t) g(t-nt_0) e^{-jm\omega_0 t} = \langle f(t), G_{m,n}(t) \rangle, \quad (4)$$

where $G_{m,n}(t) = g(t-nt_0) e^{-jm\omega_0 t}$.

WAVELET TIME-FREQUENCY REPRESENTATIONS

Given an appropriate function space, the set of functions generated by recursive applications of the *scaling function* to the *mother wavelet* constitutes a set of basis functions called a *wavelet series* [4]. The *Continuous Wavelet Transform* (CWT) is similar to the Fourier Transform (or Windowed Fourier Transform) in that both involve an inner product of an input function with a scaling function. The CWT is given as [7]

$$\begin{aligned}(T^{wav}f)(a,b) &= |a|^{-1/2} \int dt f(t) \psi^*((t-b)/a) \\ &= |a|^{-1/2} \langle f(t), \psi((t-b)/a) \rangle,\end{aligned}\tag{5}$$

Variable a , the *dilation parameter*, controls the interval of compact support, or duration, for the wavelet $\psi^{a,b} = \psi((t-b)/a)$. Variable b , the *translation parameter*, determines where in time the wavelet exists. The *Discrete Wavelet Transform* (DWT) is a collection of samples of the CWT, sampled at m incremental dilations and n incremental translations, given as [7]

$$\begin{aligned}T^{wav}_{m,n}(f) &= |a_0|^{-m/2} \int dt f(t) \psi^*(a_0^{-m}t - nb_0) \\ &= |a_0|^{-m/2} \langle f(t), \psi(a_0^{-m}t - nb_0) \rangle\end{aligned}\tag{6}$$

For a given application, the optimal basis functions, or wavelet series components, must be chosen as well as the optimal scaling and translation parameters. These selections make the application of wavelet transforms more complicated than that of Fourier transforms.

Consider a DWT implemented with unity dilation, $a_0 = 1$, and a mother wavelet that is a windowed complex exponential, $\psi^{a_0,b_0} = g(t-nt_0)e^{-jm\omega_0 t}$; in this case the DWT becomes the DFT. Using unity dilation defeats the utility of allowing for a better balance between frequency resolution at lower frequency components and time resolution at higher frequency components. However, this demonstrates that the Fourier Transform can be thought of as a special case of the Wavelet Transform.

STRUCTURE AND FUNCTION OF EARLY VISION

Natural vision filtering begins with photonic refraction through the cornea and lens. The incoming light then passes through the vitreous humor and retinal cell tissue, and is focused onto a *photoreceptor mosaic* surface. Photonic energy is converted to electronic charge in the photopigment discs of the photoreceptors (rods and cones). The photoreceptors, with the help of a layer of *horizontal* cells, spread the charge in space and time within a local neighborhood of other receptors. The spread charge and photoreceptor charge are both available at the root of the photoreceptor, at the *triad synapse*. The *bipolar* cells connect to triad synapses and presumably activate signals proportional to the difference between the photoreceptor input and the horizontal cell input. Both polarities exist, called *on*-bipolars or *off*-bipolars, which respond to light and darkness respectively [11,14,23].

The photoreceptor charge is influenced by gap junctions between adjacent photoreceptors. The response from a photoreceptor aggregate can be

modeled as a spatial-temporal Gaussian with a small variance. The input from the neighboring aggregate of horizontal cells can be modeled with a similar Gaussian with a larger variance. The differencing function results in the *difference-of-Gaussian* (DOG) filter operation, resulting in a center-surround antagonistic receptive field profile. DOG and *Laplacian-of-Gaussian* (LOG) functions have been used to model the bipolar cell output [14].

The analog charge information in the retina is funneled into information pathways as it is channeled from the mosaic plane to the optic nerve [3]. The channels that exist there are the *rod* channel, initiated by rod bipolars, the *parvocellular pathway* (PP) and the *magnocellular pathway* (MP), the latter two initiated by cone bipolars [10]. Both the PP and the MP exhibit center-surround antagonistic receptive fields. PP cones are tightly connected, responding to small receptive fields, while the MP cones are more loosely connected (together with rod inputs), responding to large receptive fields.

The MP and PP perform separate spatial band-pass filtering, provide color and intensity information, and also provide temporal response channels, as illustrated in Figure 1. A relatively high degree of acuity is achieved in each domain from these few filters. The MP is sensitive to low spatial frequencies and broad color intensities, which provide basic information of the objects in the image. The PP is known to be sensitive to higher spatial frequencies and chromatic differences, which add detail and resolution [17]. In the color domain, the PP provides color opponency and thus spectral specificity, and the MP provides color non-opponency and thus overall intensity [9,12]. In the time domain, the PP provides slowly varying dynamics, while the MP provides transient responses to image dynamics.

Figure 1. Natural Vision Information Channels.

The color opponent PP responds to spatial detail, slowly-varying image dynamics, and chromatic detail. The color non-opponent MP responds to spatial averages, rapid transients, and intensity variations.

WAVELET FEATURES INHERENT IN VISION PROCESSING

Wavelet bases can be subdivided into *orthogonal* or *nonorthogonal* and *complete* or *noncomplete* categories [21]. A set of basis functions is *orthogonal* if the inner product of any two different basis functions is zero, and *complete* if no non-zero function in the space is orthogonal to every basis vector [4,22]. Orthogonality and completeness ensure a unique transformed representation of each function within the given space. These are the general requirements for the selection of wavelet bases in compression applications.

However, biological systems are not concerned with information storage for perfect reconstruction. Any machine-vision application requiring some action to be taken based on an understanding of the image content will also fit the general description. In fact, many biological processing functions are considered to be *non-orthogonal* [8]. The task is processing information in order to take some action, not processing information for later reconstruction. The redundancy of vision filters is balanced by the need for efficiency, simplicity, and robustness. Information redundancy results in unnecessary hardware and interconnections, but often redundancy may be required to sufficiently span the information space inherent in the environment. The cost of supporting the redundancy may be less significant than the benefit of using simpler processing elements that degrade gracefully.

Wavelet Filter Banks and Vision Pathways

The MP and PP decompose the natural input image into local average and local detail components, respectively. Images can also be decomposed into wavelet components using *quadrature mirror filtering* (QMF), resulting in a series of averaging components and another series of detailing components [7,19]. QMF is a special case of *subband coding*, where filtered components represent the lower and upper frequency halves of the original signal bandwidth. If the analyzing filter coefficients are symmetric, then the synthesizing components are mirrored with respect to the half-band value, thus the term *quadrature mirror*. A variety of applications have emerged from the remarkable QMF reconstruction capabilities [13,16,18].

Vision pathways (MP and PP) and QMF filter banks both therefore break up the input image signal into high and low frequency components. Recent spectral analysis of Gaussian-derived vision models indicate a remarkable retention of information in spite of the nonorthogonal nature of vision filters [2,3].

Examples of Waveform Translation and Dilation in Vision Systems

As described previously, a wavelet basis is generated by recursively applying scaling function to a mother wavelet [4,7]. The essential characteristics of the

wavelet scaling function are the translation and dilation parameters of (5) and (6). Structural and functional subunits of vision systems also tend to replicate a mother function with scaling and translational variations. These variations are functions of other parameters such as time, gaze direction, and *eccentricity*, the distance from the center of the retina.

Photonic spreading is a function of the finite aperture, optical imperfections, and photonic wavelength [10]. A Gaussian mother wavelet of a carefully chosen variance can be chosen to model the point spread function at the retina center (zero eccentricity) at a selected visible frequency. The complete point spread function can be modeled as a two-dimensional scaling of this mother function with respect to both eccentricity and photonic frequency.

The photoreceptor mosaic includes rods and three cone types, called *L*, *M*, and *S* for "long", "medium", and "short" visible wavelength peaks in their spectral absorption curves. Sampling in the center of the retina is limited to *L* and *M* cells, providing a significantly higher spatial resolution as compared to the periphery [10,11]. For a given viewing position, a mental model of an environment is built from a set of high resolution samples from different directions. These translations of the mosaic function are the result of eye movements [5].

Due to scaling with frequency, higher frequency wavelet components typically have smaller regions of support. The sampling density of *M* and *L* cone cells is greatly reduced as a function of increasing eccentricity. This is due to increasing cell sizes and also by the introduction of rod cells and *S* cells in the outer regions. A cellular receptive field function representing the envelope of available photonic energy stimulating either *M* or *L* cones will thus include an increasing scale factor with eccentricity.

The initial electronic processing stages of early vision are characterized by DOG filtering operations of the MP and PP. The spatial extent of both the MP and PP receptive fields increase significantly with eccentricity [10]. A DOG mother function can be scaled with respect to eccentricity to properly model vision pathways at different parts of the retina.

Efficient Use of Basis Functions

A common theme among space, time, and color domains is the minimal use of basis functions. There are essentially four chromatic detector types, three temporal channels, and three spatial channels. Therefore, natural neuronal systems tend to use efficient combinations of only a few filters to accomplish a high degree of acuity in each domain. QMF signal reconstruction capability is a practical demonstration of extracting spectral detail from only two filters. The behavior of such synthetic applications may lead to a deeper understanding of natural phenomena.

ACKNOWLEDGMENTS

This effort was performed in-house at the Advanced Guidance Division of the Wright Laboratory Armament Directorate located at Eglin AFB FL. This research leverages recent doctoral research at the Department of Electrical Engineering of the FAMU-FSU College of Engineering in Tallahassee FL. Comments and collaboration are encouraged.

REFERENCES

- [1] Brooks, Geoffrey, "Comparison of vision and conventional wavelet filters," Proceedings of the SPIE, vol. 3078, no. 48, 1997.
- [2] Brooks, Geoffrey, "Biologically-inspired analog wavelet analyzers," Proc. of the SPIE, vol. 3016, no.8, 1997.
- [3] Brooks, Geoffrey, Modeling Retinal Processing using Wavelet Theory, Ph.D. Dissertation, Florida State University, 1996.
- [4] Chui, Charles, An Introduction to Wavelets, Academic Press, 1994.
- [5] Churchland, P., Ramachandran, V., and Sejnowski, T., "A critique of pure vision," in Koch, C., and Davis, J., editors, Large-Scale Neuronal Theories of the Brain, MIT Press, Cambridge, Mass., 1994.
- [6] Dagnelie, Gislin, and Massof, Robert, "Toward an Artificial Eye," IEEE Spectrum Magazine, May, 1996.
- [7] Daubechies, Ingrid, Ten Lectures on Wavelets, Society for Industrial and Applied Mathematics, Philadelphia, 1992.
- [8] Daugman, John, "Complete Discrete 2-D Gabor Transforms by Neural Networks for Image Analysis and Compression," IEEE Transactions on Acoustic, Speech, and Signal Processing, vol. 36, no. 7, July 1988.
- [9] De Valois, R. L., & De Valois, K. K., "A multi-stage color model," Vision Research, vol. 33, no. 8, 1993.
- [10] De Valois, R. L., & De Valois, K. K., Spatial Vision, Oxford University Press, New York, NY, 1988.
- [11] Dowling, John E., The Retina: An Approachable Part of the Brain, Harvard University Press, Cambridge, Mass., 1987.
- [12] Gath, S. Lee, "Model for color vision and light adaptation," Journal of the Optical Society of America, vol. 8, no. 6, 1991.
- [13] Mallat, Stephane, "A Theory for Multiresolution Signal Decomposition: The Wavelet Rep.," IEEE Transactions on PAMI, vol. 11, no. 7, 1989.
- [14] Marr, David, Vision, W. H. Freeman and Company, New York, 1982.
- [15] Oppenheim, Alan, and Schafer, Ronald, Discrete-time Signal Processing, Prentice Hall, Englewood Cliffs, NJ, 1989.
- [16] Rioul, Oliver, and Vetterli, Martin, "Wavelets and Signal Processing," IEEE Signal Processing Magazine, October, 1991.

- [17] Sherman, S. M., "Parallel W-, X-, and Y-cell pathways in the cat: a model for visual function," in Rose, D., and Dobson, V., eds, Models of the Visual Cortex, J. Wiley & Sons, New York, NY, 1985.
- [18] Strang, Gilbert, "Wavelets," American Scientist, vol. 82, May-June, 1994.
- [19] Strang, Gilbert, and Nguyen, Truong, Wavelets and Filter Banks, Wellesley-Cambridge Press, Wellesley, MA, 1996.
- [20] Strum, Robert, and Kirk, Donald, First Principles of Discrete Systems and Digital Signal Processing, Addison-Wesley, New York, 1989.
- [21] Szu, Harold, "Wavelet theory and applications," short course notes for the SPIE AeroSense95 Conference, Orlando, FL, April 1995.
- [22] Walter, Gilbert, Wavelets and Other Orthogonal Systems with Applications, CRC Press, Inc., Boca Raton, FL, 1994.
- [23] Werblin, Frank S. and Teeters, Jeffrey L., "Real time simulation of the retina allowing visualization of each processing stage," Proceedings of the SPIE vol. 1472, 1991.

^a Further author information: WL/MNGA, Eglin ABF, FL 32542 Email: brooksg@eglin.af.mil