A Fast Approximation to the Continuous Wavelet Transform with Applications*

Kathrin Berkner and Raymond O. Wells, Jr.
Department of Mathematics, Rice University
Houston, TX 77251-1892, USA
berkner@cml.rice.edu, wells@rice.edu

Abstract

We propose an approximation of a continuous wavelet transform (CWT) which is based on a hierarchical scheme, similar to the fast discrete wavelet transform. This approximation keeps redundancies in time and scale. Furthermore, it preserves properties of the CWT regarding the characterization of singularities and leads to efficient applications in detection and characterization of singularities.

1. Introduction

A generic image contains information distributed at a number of different scales. Multiresolution analysis with wavelets provides an important tool to decompose information in a signal into information at separate scales. The popular fast discrete wavelet transform (FDWT) uses orthogonality conditions to decompose an image at dyadic scales (see e.g. [3]). In particular, since the wavelets form an orthogonal basis this dyadic decomposition gives a complete representation of the image involving only $O(N)$ computations. It is known that for some applications, such as noise reduction, keeping time redundant information and using, e.g., the undecimated DWT and wavelet packets ([2,7]) yields better results in terms of smoothness, approximation error and signal-to-noise-ratio.

The largest amount of redundancy is contained in the continuous wavelet transform (CWT), which keeps not only redundant information in time, but also in scale. The CWT of a function f is given by

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

for a suitable wavelet function ψ (see, e.g. [3]). Some of the most popular wavelet functions in image analysis are derivatives of the Gaussian function $g(x) = \frac{1}{\sqrt{2\pi}} exp(-1/2x^2)$. The origins of applications of multiscale analysis of functions in wavelet domains of these wavelets go back to Witkin [13].

Influenced by the new field of wavelet analysis, Mallat and Hwang ([8]) presented a theory which shows how the CWT can be used to detect and characterize singularities of functions. In order to describe local properties of functions, most of the results in [8] are valid only for compactly supported wavelet functions. Applications of the results to signal analysis yield expensive computations in the continuous wavelet domain. In [4], three parameters are derived from parameterized curves in the continuous wavelet domain. The three parameters characterize a singularity of the signal and are successfully applied to multiscale edge detection. The experiments showed that redundant information in the wavelet representation is also useful for edge detection. In order to trace the curves, the application of nonlinear continuation methods is necessary. However, the computational complexity of $O(N^3)$ is still too high for efficient applications.

In this paper, we present a method for approximating a CWT. The method is based on binomial coefficients and it is similar to the hierarchical scheme of the FDWT. Since we use filters with linearly increasing length as discrete wavelet filters, we compute wavelet coefficients not only at dyadic scales, but at integer scales. In contrast to the FDWT, we achieve a much finer grid of wavelet coefficients. For the order of complexity we get $O(N^2)$, the same order as the methods discussed in [12], but with a much smaller constant ($C = 2$). The advantage of the CWT compared to the FDWT is the localization and characterization of singularities of functions. We show how these properties of the CWT can be transfered to the presented approximation scheme.

The paper is organized as follows. We review the properties of the CWT regarding the description of singularities in Section 2. Section 3 explains the hierarchical method and its approximating properties of the discrete wavelet functions to derivatives of the Gaussian function. Furthermore, the main results from Section 2 are transfered to the discretization. In Section 4 we compare our hierarchical method with

 Published in the *Proceedings of Asilomar'97*, November 2-5, Pacific, Grove, CA, 1997. This work was supported in part by Texas ATP and Alexander von Humboldt-Stiftung.

the time redundant DWT from [9]. The paper is concluded in Section 5.

2. Continuous wavelet transforms

It has been shown by several authors that the CWT is a powerful tool to analyze local regularity properties of functions (see, e.g. [6]). Those local regularities are, e.g., important for edge detection, since edges are modeled as singularities of functions. Singularities of functions are typically characterized by their Lipschitz exponent.

Definition 1 *A function is said to be Lipschitz α in x_0 ($f \in \Gamma^\alpha(x_0)$), if there exist constants $h_0 > 0$, $C_f > 0$ and a polynomial P_{n,x_0} of degree n, such that*

$$|f(x_0 + h) - P_{n,x_0}(h)| \leq C_f |h|^\alpha \tag{2}$$

for $|h| < h_0$ and $n < \alpha \leq n + 1$.

It follows from this definition that there are three parameters which characterize a Lipschitz singularity, namely the constants C_f and h_0 and the Lipschitz exponent α. Of course, a unique reconstruction of the function f from only these three parameters is not possible, but they provide a useful tool to distinguish between various different Lipschitz singularities.

The following theorem concerning the scaling behavior of the CWT of a function $f \in \Gamma^\alpha(x_0)$ is due to Holschneider & Tchamitchian (see e.g. [3]).

Theorem 1 *Let ψ be a wavelet with compact support $[-K, K]$ and suppose that $\int(1+|x|)|\psi(x)|dx < \infty$. Moreover, let f be a bounded function in $\Gamma^\alpha(x_0)$, $\alpha \in (0, 1]$. Then for $b + aK \subset (x_0 - h_0, x_0 + h_0)$ there exists a constant $B_\psi > 0$, such that*

$$|W_\psi[f](b, a)| \leq B_\psi \cdot C_f a^{\frac{1}{2}} (a^\alpha + |b - x_0|^\alpha). \tag{3}$$

This theorem guarantees that the singularity parameter C_f, h_0 and α have analogues in the wavelet domain: The scaling behavior of the wavelet coefficients is described by the same exponent α. The parameter h_0 has a corresponding pair (b_0, a_0) in the wavelet domain with $|b_0 + a_0 K - x_0| = h_0$, and corresponding to C_ψ is $B = B_\psi C_f$.

In [8], the authors use this result to show that Lipschitz singularities can be detected from maxima in the wavelet domain. In detail, these maxima are points where the map $b \mapsto |W_\psi[f](b, a)|$ has a local maximum in (b_0, a). They are called *modulus maximum points of $W_\psi[f]$*. Using wavelets which are derivatives of a smoothing function, Mallat & Hwang show the following result:

Theorem 2 *Let ψ be the n-th derivative of an $(n + 1)$ times differentiable smoothing function with compact support and let $f \in L^1([c, d])$. Then the closure of the set of points $x \in (c, d)$ where f is not Lipschitz of order n is included in the closure of the modulus maximum points of $W_\psi[f]$.*

This result has the consequence that a Lipschitz singularity x_0 with exponent $\alpha < n$ can be detected by modulus maximum points in the wavelet domain. If the wavelet ψ is continuously differentiable, then the modulus maximum points must form one-dimensional continua, the *modulus maxima curves*. In general, it is not possible to show for arbitrary wavelet functions and arbitrary functions f that a modulus maxima curve must evolve in a singularity of the function. However, for particular choices of wavelets or under certain conditions on f it is possible to show that every modulus maxima curve is an offspring of a singularity of f, see [1,14] for details.

Theorems 1 and 2 show how to extract information on the singularities of the function f using the CWT. It is more or less obvious that a DWT is not capable of detecting the fine structure of singularities, since the discretization is too coarse.

3. Approximation of the CWT by the scale redundant DWT

There are several methods of approximating the CWT. In [11] algorithms are presented for computing the CWT at dyadic scales. Approximations of the CWT using finer samplings under the assumption of orthogonality properties are discussed in [12].

We want to approximate wavelets which are derivatives of the Gaussian functions. Due to their simple and symmetric structure, these wavelets are very useful for edge detection. Our goal is to find a discretization that has a low complexity and keeps the important properties of the CWT mentioned in the previous section.

We consider the *binomial sequence of order N* given by

$$\rho_N(j) = \begin{cases} \frac{1}{2^N}\binom{N}{j} & , \quad 0 \leq j \leq N \\ 0 & , \quad \text{otherwise.} \end{cases} \tag{4}$$

In terms of probability theory, the sequence $\rho_N(j)$ is the well known binomial distribution. Applying the central limit theorem to the binomial distribution we get the following approximating property of the discrete function $\rho_N(j)$ to the Gaussian function g (see e.g. [10]).

Theorem 3 *Let $\phi_{\tilde{\rho}_N}$ be the characteristic function of the function $\tilde{\rho}_N([j - \frac{N}{2}]/[\sqrt(N)/2]) = \rho_N(j)$ and ϕ_g the characteristic function of the Gaussian function g, i.e. $\phi_{\tilde{\rho}_N}(\omega) = \hat{\tilde{\rho}_N}(-\omega)$ and $\phi_g(\omega) = \hat{g}(-\omega)$. Then for $\omega \in \mathbb{R}$ we have*

$$|\log \phi_g(\omega) - \log \phi_{\tilde{\rho}_N}(\omega)| = O(N^{-\frac{1}{2}}). \tag{5}$$

That means that $\phi_{\tilde{\rho}_N}$ converges pointwise to ϕ_g for $N \to \infty$.

We define the inner product of the function $\rho_N(j)$ with a sequence of function values $f(k)$ as

$$T_N[f](k) = c_{N,k} = \sum_{j\in\mathbb{Z}} \rho_N(j) f(k+j). \qquad (6)$$

Using the the relation $\binom{N}{j} + \binom{N}{j+1} = \binom{N+1}{j+1}$ for binomial coefficients we get the following recursive formula for the sequence $(c_{N,k})$:

$$c_{N+1,k} = \frac{1}{2}[c_{N,k-1} + c_{N,k}]. \qquad (7)$$

With similar arguments we obtain similar results for approximations of the r-th derivative of the Gaussian function, given by the sequence $\rho_N^r(j) = \nabla^r \rho_N(j)$ defined by

$$\nabla^r \rho_N(j) = \begin{cases} \sum_{l=0}^{r} (-1)^l \binom{r}{l} \rho_N(j-l) & , \quad 0 \le j \le N \\ 0 & , \quad \text{otherwise} \end{cases}$$

For the correlations

$$T_N^r[f](k) = c_{N,k}^r = \sum_{j\in\mathbb{Z}} \rho_N^r(j) f(k+j) \qquad (8)$$

we get the corresponding recursive formula

$$c_{N+1,k}^r = \frac{1}{2}[c_{N,k-1}^r + c_{N,k}^r]. \qquad (9)$$

If we consider $T_N^r[f](k)$ for variable N and k, we write $T^r[f](N,k)$.

The discrete functions ρ_N^r have the wavelet property, i.e., $\sum \rho_N^r(j) = 0$ for $r > 0$. Moreover, it can be shown that ρ_N^r has exactly r discrete vanishing moments, i.e. $\sum_{j\in\mathbb{Z}} j^k \rho_N^r(j) = 0$ for $k < r$ and $\sum_{j\in\mathbb{Z}} j^r \rho_N^r(j) \neq 0$ (see e.g. [5]). The functions ρ_N^r do not satisfy the orthogonality conditions as discrete orthogonal or biorthogonal wavelets. The transform given by $T^r[f]$ keeps redundant information in time and scale. Since the computation of the coefficients $c_{N,k}^r$ at integer scales N is done by a hierarchical scheme similarly to the fast DWT at dyadic scales we call this transform a *scale redundant discrete wavelet transform* (SRDWT). Since we can compute the values $c_{N+1,k}^r$ as a sum of only two values at scale N the computational complexity of our hierarchical scheme is $2N$. It follows from the approximating properties in Eq.(5) that we have the correspondence $a \sim \sqrt{N}/2$.

Applying the approximating property from Theorem 3, the result from Theorem 1 transfers to the discrete situation in the following way.

Theorem 4 *Let $f \in \Gamma^\alpha(x_0)$, $\alpha \in 0,1]$, δ be the discretization stepsize and $r > 0$. Then there exists a constant $\tilde{B} > 0$ such that*

$$|T_N^r[f](\delta k)| \le \tilde{B} ||\rho_N^r||_{l^1} (|k-x_0|^\alpha + N^\alpha) \qquad (10)$$

for $k + N - x_0 < \frac{h_0}{\delta}$. If $0 < k - x_0 < N$ there exists a constant $C^* \sim C_f$ such that

$$|T_N^r[f](\delta k)| \le C^* ||\rho_N^r||_{l^1} N^\alpha. \qquad (11)$$

The set $\{(N,k) \,|\, 0 < k - x_0 < N\}$ forms the *influence cone* of the singularity x_0. By Theorem 4 it follows that the three singularity parameters from Definition 1 have analogue parameters in the discrete domain of the SRDWT:

- The influence interval bound h_0 has an analogue *scale* N_0, namely $N_0 \sim [\frac{h_0}{\delta} + x_0]$.
- The scaling behavior of $T^r[f](N,k)$ is described by the same *Lipschitz exponent* α.
- The constant C^* is proportional to C_f. We call C^* the *derivative parameter*.

In order to characterize the regularity of a sampled sequence $f(k)$ we define *modulus maximum points of $T^r[f]$* as points (N^*, k^*) such that

$$\begin{array}{lll} c_{N,k}^r > 0 & \text{and} & c_{N,k-1}^r < c_{N,k}^r > c_{N,k+1}^r, \quad \text{or} \\ c_{N,k}^r < 0 & \text{and} & c_{N,k-1}^r > c_{N,k}^r < c_{N,k+1}^r. \end{array}$$

We can show a property of modulus maximum points of $T^r[f]$ which is similar to the continuous situation, where modulus maximum points of $W_{g^{(r)}}[f]$ form curves which go down to the lowest scale.

Lemma 1 *Let (N^*, k^*) be a modulus maximum point of $T^r[f]$. Then there exists $N_{max} > 0$ and a sequence $\{(m, k_m)\}$, $m = 1, \dots N_{max}$, of modulus maximum points of $T^r[f]$ with $(N^*, k^*) \in \{(m, k_m)\}$ such that:*

- *The sign of c_{m,k_m} does not change for $m = 1, \dots N_{max}$,*
- $k_{m+1} \in \{k_m, k_m + 1\}$.

The sequence $\{(m, k_m)\}$ is called a modulus maxima sequence of $T^r[f]$. *The point $(1, k_1)$ is called the* origin *of the modulus maxima sequence.*

This lemma leads to a simple procedure of tracking modulus maxima sequences of $T^r[f]$ from scale $N = 1$ to larges scale. We do not need to apply numerical continuation methods as in [4]. The localization of changes of the regularity of the function f can now be formulated in the following way.

Lemma 2 *Let $f(k)$ be a sequence of sampled function values of $f \in L^2(\mathbb{R})$. Then $(1,k)$ is an origin of a modulus maxima sequence of $T^r[f]$ iff*

$$\begin{array}{lll} \nabla^r f(k) > 0 & : & \nabla^r f(k) - \nabla^r f(k-1) > 0, \\ & & \nabla^r f(k) - \nabla^r f(k+1) > 0 \quad \text{or} \\ \nabla^r f(k) < 0 & : & \nabla^r f(k) - \nabla^r f(k-1) < 0, \\ & & \nabla^r f(k) - \nabla^r f(k+1) < 0. \end{array}$$

Lemma 1 guarantees that a modulus maximum sequence always lies inside the influence cone of a singularity. Therefore, Eq.(11) reduces to the following:

$$|c^r_{N,k_N}| \leq \tilde{B^*} \parallel \rho^r_N \parallel_{l_1} N^\alpha. \tag{12}$$

As a consequence, we can estimate the Lipschitz exponent from the scaling behavior of the wavelet coefficients on a modulus maxima sequence. Numerical examples follow in the next section.

4. Examples

Since sharp edges correspond to singularities with Lipschitz exponent $\alpha < 1$, it is sufficient for edge detection to consider the case $r = 1$.

In order to measure the three parameters α, C^* and N_0 from the previous section, we have to estimate the norm $||\rho^r_N||_{l^2}$ from Theorem 4.

Lemma 3 *Let $N = 2p$ for $p \in \mathbb{N}$. Then*

$$||\rho^1_N||_{l^1} = \binom{N}{\frac{N}{2}} \sim N^{-\frac{1}{2}}. \tag{13}$$

Figure 1(a) shows a function with five singularities $x_1, \ldots x_5$ and different Hölder exponents $\alpha(x_1) = 0.3, \alpha(x_2) = 0.3, \alpha(x_3) = 0.5, \alpha(x_4) = 0.7, \alpha(x_5) = 0.5$. The localization of singularities by modulus maxima sequences of $T^1[f]$ computed with the SRDW is shown in Figure 1(b). In order to estimate the different Hölder exponents, we have to estimate the slope of the linear graph in the double-logarithmic presentation log $|c^1_{N,k_N}|$ vs $\log N$. We can see in Figure 1(c) that the linear behavior breaks down at a scale N_0. We use this scale as an estimation for the influence interval characteristic h_0. The derivative parameter C^* can be estimated by the y-intersept in the double-logarithmic plot. A description of a singularity by the three parameters (computed from modulus maxima sequences) makes it possible to distinguish between the singularities.

In comparison to the analysis of singularities with the SRDWT we show the wavelet transform of the signal from Figure 1(a) at various scales using the method presented in [8,9]. That method computes an undecimated discrete wavelet transform at dyadic scales using an approximation of the first derivative of a cubic spline as the analyzing wavelet. The disadvantages of this method compared to the SRDWT are the following. In general, there is no relation between modulus maximum points at various scales. The wavelet transform does not guarantee that extrema at one dyadic scale correspond to extrema at the next lower scale. Moreover, some extrema which correspond to a singularity "disappear" between two dyadic scales (Figure 2: x_3 disappears between scale 8 and 16). This only gives a weak estimation of the influence interval. Based on these facts, a characterization of singularities by Lipschitz exponents and influence scale is, in general, not possible when using a time redundant DWT at dyadic scales. Since the FDWT does not even keep redundant information in time as the method presented in [8,9], the singularity characterization does certainly not work with this method.

5. Conclusions

We have introduced an approximation to the CWT using a hierachical scheme which needs only 2N summations per scale. Our discretization keeps the important property of the CWT concerning the characterization of singularities. We have shown how the theoretical results from the CWT are transfered to the discrete situation and have illustrated them in an example.

References

[1] Berkner, K., Wells Jr, R.O. A geometric description of singularities in the wavelet domain. in preparation.

[2] Coifman, R.R., Donoho, D.L. Translation invariant denoising. In A. Antoniades, editor, *Wavelets and Statistics, Springer Lecture Notes*. Springer-Verlag, 1995.

[3] Daubechies, I. *Ten Lectures on Wavelets*. SIAM Publ., 1992.

[4] Evertsz, C.J.G., Berkner, K., Berghorn, W. A local multiscale characterization of edges applying the wavelet transform. In *Proc. Nato A.S.I., Fractal Image Encoding and Analysis, Trondheim, July 1995*, 1997.

[5] Feller, W. *An Introduction to Probability Theory and Its Applications*. John Wiley & Sons, Inc., 2. edition, 1962.

[6] Holschneider, M. *Wavelets - An Analysis Tool*. Oxford Academic Press, 1995.

[7] Lang, M. et al. Noise reduction using an undecimated discrete wavelet transform. *IEEE Signal Processing Letters*, 3:10–12, 1996.

[8] Mallat, S., Hwang, W.L. Singularity detection and processing with wavelets. *IEEE Trans. Inform. Theor.*, 38(2):617–643, 1992.

[9] Mallat, S., Zhong, S. Characterization of signals from multiscale edges. *IEEE Trans. Patt. Anal. Mach. Intel.*, 14(7):710–732, 1992.

[10] Papoulis, A. *Probability, Random Variables and Stochastic Processes*. McGraw-Hill, 1965.

[11] Shensa, M.J. The discrete wavelet transform: wedding the à trous and Mallat algorithm. *IEEE Trans. Sig. Proc.*, 40(10):2464–2482, 1992.

[12] Vrhel, M.J., Lee, C., Unser, M. Comparison of algorithms for the fast computation of the continuous wavelet transform. In *Proc. SPIE, Denver, CO*, volume 2825, pages 422–431, 1996.

[13] Witkin, A. Scale-space filtering. In *Proc. Int. Joint Conf. Artif. Intell., Karlsruhe*, pages 1019–1021, 1983.

[14] Yuille, A.L., Poggio, T. Scaling theorems for zero-crossing. *IEEE Trans. Patt. Anal. Mach. Intel.*, 8(1):15–25, 1986.

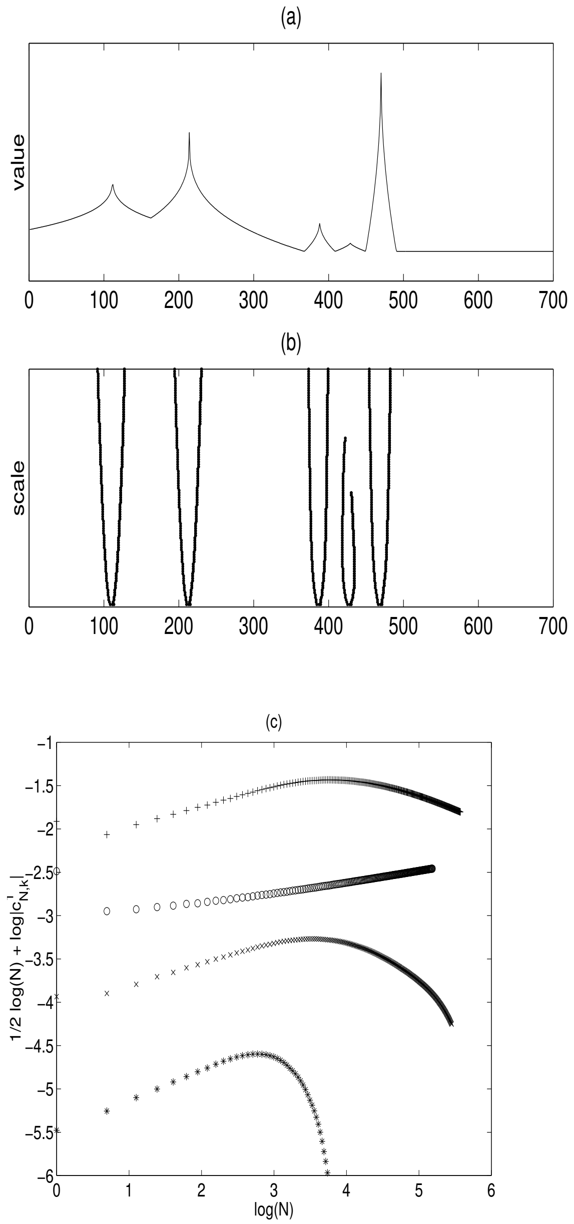

Figure 1. Wavelet analysis of singularities with different Lipschitz exponents. *Top to bottom:* **(a) Signal with 5 singularities (***left to right***: $x_1, \dots x_5$). (b) Modulus maxima sequences of the SRDWT. (c) Loglog plots of the values $N^{1/2}|c^1_{N,k}|$ on modulus maxima sequences from (b): 'o' (x_2), 'x' (x_3), '$*$' (x_4), '+' (x_5).**

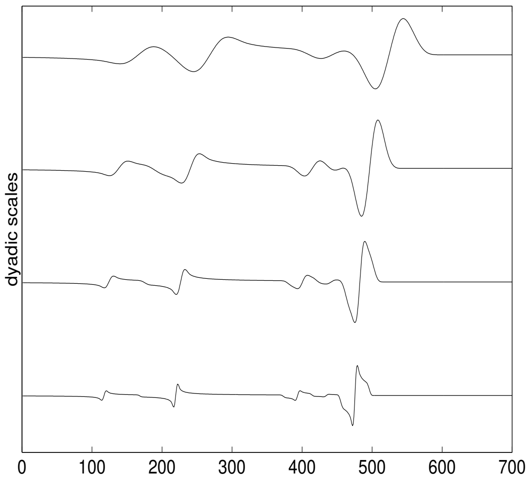

Figure 2. Wavelet transform with the derivative of a cubic spline at dyadic scales $2, 4, 8, 16$.