CWI Quarterly Volume 11 (4) 1998, pp. 323 – 345

The Continuous Wavelet Transform in Image Processing

J-P. Antoine
Institut de Physique Théorique
Université Catholique de Louvain
B–1348 Louvain-la-Neuve, Belgium
e-mail: antoine@fyma.ucl.ac.be

We begin with a short review of the 2-D continuous wavelet transform (CWT), covering successively the mathematical properties, the various analyzing wavelets and the calibration problem. Then, after a short outline of the discrete wavelet transform, we present a fast algorithm for the CWT. Then we discuss the group-theoretical background of the CWT, which allows a straightforward extension to more general situations. We conclude with some physical applications.

1. Introduction: The Continuous Wavelet Transform

The one-dimensional wavelet transform [12, 19, 32] has found nowadays many applications to various fields of physics, mathematics and signal processing [16, 33, 34]. The original motivation was to design a method of analysis suitable for nonstationary, highly inhomogeneous signals (such as speech), for which Fourier analysis is inadequate. The outcome is a *time-scale* analysis, based on the wavelet transform (WT):

$$S(b,a) = a^{-1/2} \int \overline{\psi\left(a^{-1}(t-b)\right)}\, s(t)\, dt \equiv \langle \psi_{b,a} | s \rangle, \tag{1}$$

where $a > 0$ is a scale parameter and $b \in \mathbb{R}$ a translation parameter. In relation (1), s is a finite energy signal, the function ψ, the analyzing wavelet, is assumed to be well localized *both* in the time domain and in the frequency domain, and the bracket denotes the usual scalar product in $L^2(\mathbb{R}, dt)$. In addition ψ must satisfy an admissibility condition, which in most cases may be reduced to the requirement that ψ has zero mean (hence it is sufficiently oscillating): $\int \psi(t)\, dt = 0$. Combining this condition with the localization properties of $\psi(t)$ and its Fourier transform $\widehat{\psi}(\omega)$, one sees that the WT $s \mapsto S$ provides a *local*

filtering, both in time (b) and in scale (a), which works at constant relative bandwidth, $\Delta\omega/\omega =$ constant. Thus it is more efficient at high frequency, i.e. small scales, in particular for the detection of singularities in the signal. In addition, the transformation $s(x) \mapsto S(a,b)$ may be inverted exactly and yields a reconstruction formula, which amounts to a decomposition of the signal in terms of dilated, translated copies $\psi_{b,a}$ of the basic wavelet ψ.

Of course, the numerical implementation requires the discretization of integrals. In particular, the reconstruction formula expresses the signal as a linear superposition of a discrete family $\{\psi_{b_i,a_j}\}$. However, in general, this approach does *not* lead to an orthonormal basis. In order to achieve this, it is necessary to exploit a totally different approach, based on *multiresolution analysis* [19, 28, 29, 32]. The resulting discrete wavelet transform (DWT) has clear advantages in certain problems, such as data compression.

Quite naturally, the DWT extends to 2-D and it has become a standard tool in image processing [19, 28]. It proves indeed quite efficient, yielding for instance excellent compression rates . For that purpose, the 2-D multiresolution analysis is usually taken as the tensor product of two 1-D analyses, one in x and one in y. It is thus clearly bound to the Cartesian geometry, and this is a distinct disadvantage when it comes to detecting directions, although it is natural in other instances (TV, for instance).

Fortunately, the continuous wavelet transform (CWT) may also be extended to 2 (or more) dimensions [35], with exactly the same properties as in the 1-D case. Here again the mechanism of the WT is easily understood from its very definition as a convolution (a detailed discussion is given in Section 2):

$$S(\vec{b}, a, \theta) \sim \int \overline{\psi\left(a^{-1} r_{-\theta}\left(\vec{x} - \vec{b}\right)\right)} s(\vec{x})\, d^2\vec{x}, \tag{2}$$

where s is the signal and ψ is the analyzing wavelet, which is translated by $\vec{b} \in \mathbb{R}^2$, dilated by $a > 0$ and rotated by an angle θ ($r_{-\theta}$ is the rotation operator). Since the wavelet ψ is required to have zero mean, we have again a filtering effect, i.e. the analysis is *local* in all four parameters $\vec{b}, a, \theta$, and here too it is particularly efficient at detecting discontinuities in images. When compared to the 1-D case, the new fact here is the presence of the rotation degree of freedom. This is crucial for detecting *oriented features* of the signal, that is, regions where the amplitude is regular along one direction and has a sharp variation along the perpendicular direction, for instance, edges or contours [25]. The CWT is a very efficient tool in this respect, provided one uses a *directional* wavelet, that is, a wavelet which has itself an intrinsic orientation (for instance, it contains a plane wave).

It is a quite common opinion that the CWT is too time consuming for any practical use in image processing. This is, we think, a misconception. Not only is it better adapted in a number of situations, but in addition fast algorithms have been designed recently that make it truly competitive numerically [24].

The aim of this paper is to survey the theory and some applications of the 2-D continuous WT. We begin in Section 2 by briefly reviewing the mathematical

basis of the theory and the problem of practical implementation. Then, in Section 3 and 4, respectively, we describe various types of 2-D wavelets and discuss the problem of calibration. In Section 5, we compare the CWT with the discrete approach, and present fast algorithms for the CWT. In Section 6, finally, we review some applications of the CWT to different physical problems.

2. The continuous WT in two dimensions

2.1. Construction and main properties

We begin by reviewing briefly the basic properties of the CWT in 2 dimensions, which are completely analogous to those familiar in the 1-D case.

We consider 2-D signals of finite energy, represented by complex-valued, square integrable functions $s \in L^2(\mathbb{R}^2, d^2\vec{x})$. This condition may be relaxed, to allow, for instance, a plane wave or a δ function. In practice, a black and white image will be represented by a bounded non-negative function: $0 \leq s(\vec{x}) \leq M$, $\forall\, \vec{x} \in \mathbb{R}^2$ $(M > 0)$, the discrete values of $s(\vec{x})$ corresponding to the level of gray of each pixel.

Given a signal $s \in L^2(\mathbb{R}^2, d^2\vec{x})$, we may transform it by translation, rotation and global dilation [35]. This gives, in position and momentum (or spatial frequency) space, respectively:

$$s_{\vec{b},a,\theta}(\vec{x}) \;=\; a^{-1} s\left(a^{-1} r_{-\theta}\left(\vec{x} - \vec{b}\right)\right), \tag{3}$$

$$\widehat{s_{\vec{b},a,\theta}}(\vec{k}) \;=\; a\, e^{-i\vec{b}.\vec{k}}\, \widehat{s}(a r_{-\theta}(\vec{k})). \tag{4}$$

In these relations, $\vec{b} \in \mathbb{R}^2$ is the translation parameter, $a > 0$ the dilation, and $r_{-\theta}(0 \leq \theta < 2\pi)$ denotes the familiar 2×2 rotation matrix. As usual, the hat denotes a 2-D Fourier transform. Clearly, the correspondence $s \mapsto s_{\vec{b},a,\theta}$ is a unitary map.

By definition, a *wavelet* is a complex-valued function $\psi \in L^2(\mathbb{R}^2, d^2\vec{x})$ satisfying the admissibility condition

$$c_\psi \equiv (2\pi)^2 \int |\widehat{\psi}(\vec{k})|^2\, \frac{d^2\vec{k}}{|\vec{k}|^2} < \infty. \tag{5}$$

If ψ is regular enough, the admissibility condition simply means that the wavelet has zero mean:

$$\widehat{\psi}(\vec{0}) = 0 \iff \int \psi(\vec{x})\, d^2\vec{x} = 0. \tag{6}$$

Clearly the map $s \mapsto s_{\vec{b},a,\theta}$ preserves the admissibility condition (5). Thus the given wavelet ψ generates, by translation, rotation or dilation, the whole family $\{\psi_{\vec{b},a,\theta}\}$, indexed by $a > 0, \theta \in [0, 2\pi), \vec{b} \in \mathbb{R}^2$. It is easily seen that the linear span of this family is dense in $L^2(\mathbb{R}^2)$. In the sequel we will denote by G this 4-dimensional parameter space. Indeed, as this notation suggests,

the whole construction has a group-theoretical backbone. The parameter space G is in fact the so-called *similitude group* of the plane, composed precisely of translations, rotations and dilations, and the 2-D CWT may be derived from a unitary representation of it. We shall come back to this point in Section 6.

Given a signal $s \in L^2(\mathbb{R}^2)$, its CWT with respect to the wavelet ψ is:

$$S(\vec{b}, a, \theta) \;=\; \langle \psi_{\vec{b},a,\theta} | s \rangle \;=\; a^{-1} \int \overline{\psi(a^{-1} r_{-\theta}(\vec{x} - \vec{b}))}\; s(\vec{x})\; d^2\vec{x} \tag{7}$$

$$\;=\; a \int e^{i\vec{b}.\vec{k}}\; \overline{\widehat{\psi}(a r_{-\theta}(\vec{k}))}\; \widehat{s}(\vec{k})\; d^2\vec{k}. \tag{8}$$

The properties of the wavelet transform are best expressed in terms of the linear map $W_\psi : s \mapsto c_\psi^{-1/2} S$. They may be summarized as follows:

(i) W_ψ is *covariant* under translations, dilations and rotations. For instance:

$$S_{\vec{b}_o}(\vec{b}, a, \theta) = S(\vec{b} - \vec{b}_o, a, \theta), \tag{9}$$

where $S_{\vec{b}_o}$ is the transform of $s_{\vec{b}_o}(\vec{x}) = s(\vec{x} - \vec{b}_o)$.

(ii) W_ψ *conserves norms*:

$$c_\psi^{-1} \iiint |S(\vec{b}, a, \theta)|^2\; a^{-3} d^2\vec{b}\, da\, d\theta = \int |s(\vec{x})|^2\; d^2\vec{x}, \tag{10}$$

i.e., W_ψ is an isometry from the space of signals into the space of transforms, which is a closed subspace of $L^2(G, dg)$, where $dg \equiv a^{-3} d^2\vec{b}\, da\, d\theta$ is the natural (Haar) measure on G.

(iii) As a consequence, the map W_ψ is *invertible* on its range, and the inverse transformation is simply the adjoint of W_ψ. This means that the signal $s(\vec{x})$ may be reconstructed exactly from its transform $S(\vec{b}, a, \theta)$:

$$s(\vec{x}) = c_\psi^{-1} \iiint \psi_{\vec{b},a,\theta}(\vec{x})\; S(\vec{b}, a, \theta)\, a^{-3} d^2\vec{b}\, da\, d\theta. \tag{11}$$

In other words, the 2-D wavelet transform provides a decomposition of the signal in terms of the analyzing wavelets $\psi_{\vec{b},a,\theta}$, with coefficients $S(\vec{b}, a, \theta)$.

(iv) The projection from $L^2(G, dg)$ onto the range of W_ψ, the space of wavelet transforms, is an integral operator whose kernel $K(\vec{b}', a', \theta' | \vec{b}, a, \theta)$ is the autocorrelation function of ψ, also called *reproducing kernel:*

$$K(\vec{b}', a', \theta' | \vec{b}, a, \theta) = c_\psi^{-1}\; \langle \psi_{\vec{b}',a',\theta'} | \psi_{\vec{b},a,\theta} \rangle. \tag{12}$$

Therefore, a function $f \in L^2(G, dg)$ is the wavelet transform of a certain signal iff it satisfies the reproduction property:

$$f(\vec{b}', a', \theta') = \iiint K(\vec{b}', a', \theta' | \vec{b}, a, \theta)\; f(\vec{b}, a, \theta)\, a^{-3}\, d^2\vec{b}\, da\, d\theta. \tag{13}$$

In practice, one often uses a simplified reconstruction, instead of (11), which requires integration over the full group. For instance, if the wavelet is rotation invariant, the θ dependence drops out. More generally, one may use for reconstruction a wavelet different from the analyzing wavelet, provided the two satisfy a cross-compatibility condition. An extreme case is that of a delta distribution, with support either at the origin, or on a line. In the first case, one reconstructs the signal by resumming only over scales and angles:

$$s(\vec{x}) \sim \iint S(\vec{x}, a, \theta)\ a^{-2}\, da\, d\theta. \tag{14}$$

The second case yields a wavelet version of the inverse Radon transform, which is the mathematical basis of tomography. Further details may be found, for instance, in [26, 27, 41].

2.2. Interpretation of the CWT as a singularity scanner

In order to get a physical interpretation of the CWT, we notice that in signal analysis, as in classical electromagnetism, the L^2 norm is interpreted as the total energy of the signal. Therefore, the relation (10) suggests to interpret $|S(\vec{b}, a, \theta)|^2$ as the energy density in the wavelet parameter space.

Assume now, as in 1-D, that the wavelet ψ is fairly well localized both in position space ($\vec{x}$) and in spatial frequency space ($\vec{k}$). Then so does the transformed wavelet $\psi_{\vec{b},a,\theta}$, with effective support suitably translated by $\vec{b}$, rotated by θ and dilated by a. Because (7) is essentially a convolution with a function ψ of zero mean, the transform $S(\vec{b}, a, \theta)$ is appreciable only in those regions of parameter space $(\vec{b}, a, \theta)$ where the signal is: we get an appreciable value of S only where the wavelet $\psi_{\vec{b},a,\theta}$ 'matches' the features of the signal s. In other words, the CWT acts on a signal as a *local* filter in all 4 variables $\vec{b}, a, \theta$: $S(\vec{b}, a, \theta)$ 'sees' only that portion of the signal that 'lives' around $\vec{b}, a, \theta$ and filters out the rest. Therefore, if the wavelet is well localized, the energy density of the transform will be concentrated on the significant parts of the signal. This is the key to all the approximation schemes that make wavelets such an efficient tool.

Let us make more precise the support properties of ψ. Assume ψ and $\widehat{\psi}$ to be as well localized as possible (compatible with the Fourier uncertainty property), that is, ψ has for essential support (i.e. the region outside of which the function is numerically negligible) a 'disk' of diameter T, centered around $\vec{0}$, while $\widehat{\psi}$ has for essential support a 'disk' of diameter Ω, centered around $\vec{k}_o$. Then, for the transformed wavelets $\psi_{\vec{b},a,\theta}$ and $\widehat{\psi}_{\vec{b},a,\theta}$ we have, respectively:

(i) ess supp $\psi_{\vec{b},a,\theta}$ is a 'disk' of diameter $\simeq aT$ around $\vec{b}$, rotated by r_θ;

(ii) ess supp $\widehat{\psi}_{\vec{b},a,\theta}$ is a 'disk' of diameter $\simeq \Omega/a$ around $\vec{k}_o/a$, rotated by r_θ.

Notice that the product of the two diameters is constant. Thus the wavelet analysis operates at constant *relative bandwidth*, $\Delta k/k =$ const, where $k \equiv |\vec{k}|$. Therefore, the analysis is most efficient at high frequencies or small scales, and so it is particularly apt at detecting *discontinuities* in images, either point singularities (contours, corners) or directional features (edges, segments).

In addition to its localization properties, the wavelet ψ is often required to have a certain number of vanishing moments. This condition determines the capacity of the WT to detect singularities. Indeed, if ψ has all its moments vanishing up to order $n \geq 1$ (by the admissibility condition (6), the moment of order 0 must always vanish),

$$\int x^\alpha \, y^\beta \, \psi(\vec{x}) \, d^2\vec{x} = 0, \quad 0 \leq \alpha + \beta \leq n, \tag{15}$$

then the CWT will filter out any polynomial behavior up to degree n. For instance, the case $n = 1$ means that the CWT is blind to any linear behavior, i.e. it erases any linear trend in the signal. In general, if the wavelet has vanishing moments, the smoother parts of the signal will have very small wavelet coefficients, whereas sharp, non-stationary behavior will give rise to local maxima of the modulus of $S(\vec{b}, a, \theta)$. This explains why wavelet analysis is particularly efficient for the detection of discontinuities.

2.3. Practical implementation: The various representations

The first problem one faces in practice is that of visualization. Indeed $S(\vec{b}, a, \theta)$ is a function of four variables: two position variables $\vec{b} = (b_x, b_y) \in \mathbb{R}^2$, and the pair $(a, \theta) \in \mathbb{R}_*^+ \times [0, 2\pi)$.

In the 1-D case [18, 29], a^{-1} defines the frequency scale, thus the full parameter space of the 1-D WT, the time-scale half plane, is in fact a phase space, in the sense of Hamiltonian mechanics. Exactly the same situation prevails in 2-D: the pair (a^{-1}, θ) plays the role of spatial frequency (or momentum), expressed in polar coordinates, and so the full 4-dimensional parameter space of the 2-D WT may be interpreted as a phase space. This interpretation, which actually extends to higher dimensions, is borne out by mathematical analysis, using the group-theoretical approach (one computes the coadjoint orbits of the similitude group) [3, 4].

Now, to compute and visualize the full CWT in all four variables is hardly possible. Therefore, in order to obtain a manageable tool, some of the variables, a, θ, b_x, b_y must be fixed. In other words, one must restrict oneself to a *section* of the parameter space. There are six possible choices of two-dimensional sections, but the geometrical considerations made above indicate that two of them are more natural: either (a, θ) or (b_x, b_y) are fixed, and the WT is treated as a function of the two remaining variables. The corresponding representations have the following characteristics [3].

(1) The *position representation:* a and θ are fixed and the CWT is considered as a function of position $\vec{b}$ alone.

(2) The *scale-angle representation:* for fixed $\vec{b}$, the CWT is considered as a function of scale a and angle θ, i.e. of spatial frequency.

The position representation is the standard one, and it is useful for the general purposes of image processing: detection of position, shape and contours of objects; pattern recognition; image filtering by resynthesis after elimination of unwanted features (for instance, noise). The scale-angle representation will be particularly interesting whenever scaling behavior (as in fractals) or angular selection is important, in particular when directional wavelets are used. In fact, both representations are needed for a full understanding of the properties of the CWT in all four variables.

For the numerical evaluation, in particular for exploiting the reconstruction formula (11), one has to discretize the WT. In either representation, a systematic use of the FFT algorithm will lead to a numerical complexity of the order of $CN_1N_2(\log_2(N_1N_2))^2$, where N_1, N_2 denote the number of sampling points in the variables (b_x, b_y) or (a, θ) and C is a constant. In the former case, the geometry is Cartesian and a square lattice will give an adequate sampling grid. In the latter, the representation is in polar coordinates, and the discretization must naturally be logarithmic in the scale variable a and linear in the angle θ.

3. Choice of the analyzing wavelet

The next step is to choose an analyzing wavelet ψ. At this point, there are two possibilities, depending on the problem at hand, namely isotropic or directional wavelets.

3.1. Isotropic wavelets

If one wants to perform a pointwise analysis, that is, when no oriented features are present or relevant in the signal, one may choose an analyzing wavelet ψ which is invariant under rotation. Then the θ dependence drops out, for instance, in the reconstruction formula (11). Typical examples are:

The isotropic 2-D mexican hat or Marr wavelet : This is simply the Laplacian of a Gaussian:

$$\psi_H(\vec{x}) = (2 - |\vec{x}|^2)\, \exp(-\tfrac{1}{2}|\vec{x}|^2) = -\Delta\, \exp(-\tfrac{1}{2}|\vec{x}|^2) \tag{16}$$

(Δ denotes the Laplacian operator). This is a real, rotation invariant wavelet, originally introduced by [30]. The mexican hat is efficient for a fine pointwise analysis, but not for detecting directions. On the other hand, one may also use higher order Laplacians of the Gaussian,

$$\psi_H^{(n)}(\vec{x}) = (-\Delta)^n\, \exp(-\tfrac{1}{2}|\vec{x}|^2). \tag{17}$$

For increasing n, these wavelets have more and more vanishing moments, and are thus sensitive to increasingly sharper details. An interesting technique,

pioneered in 1-D by A. ARNÉODO [9], is to analyze the same signal with several wavelets $\psi_H^{(n)}$, for different n. The features common to all the transforms surely belong to the signal, they are not artifacts of the analysis.

Difference wavelets : Many other wavelets (or filters) have been proposed in the literature, often designed for a specific problem. An interesting class consists of wavelets obtained as the difference of two positive functions, for instance a single function h and a contracted version of the latter. If h is a smooth non-negative function, integrable and square integrable, with all moments of order one vanishing at the origin, then the function ψ given by the relation :

$$\psi(\vec{x}) = \alpha^{-2}\, h(\alpha^{-1}\vec{x}) - h(\vec{x}) \quad (0 < \alpha < 1) \tag{18}$$

is easily seen to be a wavelet satisfying the admissibility condition (6).

A typical example is the 'Difference-of-Gaussians' or DOG wavelet, obtained by taking for h a Gaussian

$$\psi_D(\vec{x}) = \alpha^{-2}\, e^{-|\vec{x}|^2/2\alpha^2} - e^{-|\vec{x}|^2/2}, \quad (0 < \alpha < 1). \tag{19}$$

The DOG filter is a good substitute for the mexican hat (for $\alpha^{-1} = 1.6$, their shapes are extremely similar), frequently used in psychophysics works [20].

3.2. Directional wavelets

When the aim is to detect oriented features (segments, edges, vector field,...) in an image, or to perform directional filtering, one has to use a wavelet which is sensitive to directions. The best angular selectivity will be obtained if ψ is *directional*, which means that the effective support of its Fourier transform $\widehat{\psi}$ is contained in a convex cone in spatial frequency space $\{\vec{k}\}$, with apex at the origin. A review of directional wavelets and their use may be found in [5]. This definition is justified by the formula (8), which says that the wavelet acts as a filter in $\vec{k}$-space (multiplication by the function $\widehat{\psi}$). Suppose the signal $s(\vec{x})$ is strongly oriented, for instance along the x-axis. Then its Fourier transform $\widehat{s}(\vec{k})$ is strongly peaked along the k_y-axis. In order to detect such a signal, with a good directional selectivity, one needs a wavelet ψ supported in a narrow cone in $\vec{k}$-space. Then the WT is negligible unless $\widehat{\psi}(\vec{k})$ is essentially aligned onto $\widehat{s}(\vec{k})$: directional selectivity demands to restrict the support of $\widehat{\psi}$, not ψ.

The 2-D Morlet wavelet : This is the prototype of a directional wavelet:

$$\psi_M(\vec{x}) \;=\; \exp(i\vec{k}_o \cdot \vec{x})\, \exp(-\tfrac{1}{2}|A\vec{x}|^2), \tag{20}$$

$$\widehat{\psi}_M(\vec{k}) \;=\; \sqrt{\epsilon}\, \exp(-\tfrac{1}{2}|A^{-1}(\vec{k} - \vec{k}_o)|^2), \tag{21}$$

The parameter $\vec{k}_o$ is the wave vector, and $A = \mathrm{diag}[\epsilon^{-1/2}, 1], \epsilon \geq 1$, is a 2×2 anisotropy matrix. We have dropped the correction term necessary to enforce

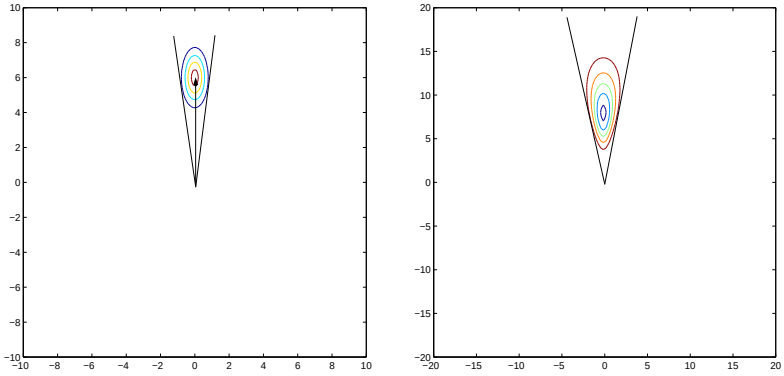

FIGURE 1. Two 2-D directional wavelets in spatial frequency space, in level curves: (a) the Morlet wavelet, with $\epsilon = 5, \vec{k}_o = (0, 6)$; (b) the Cauchy wavelet $\psi_{44}^{(\mathcal{C})}$ for $\mathcal{C} = \mathcal{C}(-10^o, 10^o)$, rotated by 90^o for the sake of comparison.

the admissibility condition $\widehat{\psi}_M(\vec{0}) = 0$, because it is numerically negligible for $|\vec{k}_o| \geq 5.6$ [19]. The wavelet ψ_M smoothes the signal in all directions, but detects the sharp transitions in the direction perpendicular to $\vec{k}_o$. In Fourier space, the effective support of the function $\widehat{\psi}_M$ is an ellipse centered at $\vec{k}_o$ and elongated in the k_y direction, thus contained in a convex cone, that becomes narrower as ϵ increases (see Figure 1 a).

Cauchy wavelets : A typical example of directional wavelets is the family of Cauchy wavelets, described in [4, 5, 6]. Let $\mathcal{C} \equiv \mathcal{C}(-\alpha, \alpha) = \{\vec{k} \in \mathbb{R}^2 \mid -\alpha \leq \arg \vec{k} \leq \alpha\}$ be the convex cone determined by the unit vectors $\vec{e}_{-\alpha}, \vec{e}_{\alpha}$. The dual cone, also convex, is $\tilde{\mathcal{C}}(-\tilde{\alpha}, \tilde{\alpha}) = \{\vec{k} \in \mathbb{R}^2 \mid \vec{k} \cdot \vec{k}' > 0, \ \forall \vec{k}' \in \mathcal{C}(-\alpha, \alpha)\}$, where $\tilde{\alpha} = -\alpha + \pi/2$, and therefore $\vec{e}_{-\alpha} \cdot \vec{e}_{\tilde{\alpha}} = \vec{e}_{\alpha} \cdot \vec{e}_{-\tilde{\alpha}} = 0$. Given the fixed vector $\vec{\eta} = (\eta, 0), \eta > 0$ (thus $\vec{\eta} \in \tilde{\mathcal{C}}$), we define the Cauchy wavelet in spatial frequency variables (see Figure 1 b):

$$\widehat{\psi}_{lm}^{(\mathcal{C})}(\vec{k}) = \begin{cases} (\vec{k} \cdot \vec{e}_{\tilde{\alpha}})^l \ (\vec{k} \cdot \vec{e}_{-\tilde{\alpha}})^m \ e^{-\vec{k} \cdot \vec{\eta}}, & \vec{k} \in \mathcal{C}(-\alpha, \alpha) \\ 0, \quad \text{otherwise.} \end{cases} \tag{22}$$

The parameters $l, m \in \mathbb{N}^*$ give the number of vanishing moments on the edges of the cone. An explicit calculation yields the following result:

$$\psi_{lm}^{(\mathcal{C})}(\vec{x}) = \text{const.} \ (\vec{z} \cdot \vec{e}_{\alpha})^{-l-1} \ (\vec{z} \cdot \vec{e}_{-\alpha})^{-m-1}, \tag{23}$$

where we have introduced the complex variable $\vec{z} = \vec{x} + i\vec{\eta} \in \mathbb{R}^2 + i\widetilde{\mathcal{C}}$. The construction generalizes in a straightforward way to any convex cone $\mathcal{C}(\alpha, \beta)$ [4, 5, 6]. In addition, if one lets $\vec{\eta}$ vary in the dual cone $\tilde{\mathcal{C}}(\tilde{\alpha}, \tilde{\beta})$, then the wavelet $\psi_{lm}^{(\mathcal{C})}(\vec{x})$ is the boundary value of a function $\psi_{lm}^{(\mathcal{C})}(\vec{z})$, holomorphic in the tube $\mathbb{R}^2 + i\widetilde{\mathcal{C}}$. This follows from general theorems [36], since the function $\widehat{\psi}_{lm}(\vec{k})$ has support in the convex cone $\mathcal{C} = \mathcal{C}(\alpha, \beta)$ and is of fast decrease at infinity.

4. Evaluation of the performances of the CWT

Given a wavelet, what is its angular and scale selectivity (resolving power)? What is the minimal sampling grid for the reconstruction formula (11) that guarantees that no information is lost? The answer to both questions resides in a *quantitative* knowledge of the properties of the wavelet, that is, the tool must be *calibrated*.

To that effect, one takes the WT of particular, standard signals. Three such tests are useful, and in each case the outcome may be viewed either at fixed (a, θ) (position representation) or at fixed $\vec{b}$ (scale-angle representation).

• *Point signal:* for a snapshot at the wavelet itself, one takes as the signal a delta function, i.e. one evaluates the impulse response of the filter:

$$\langle \psi_{\vec{b},a,\theta} | \delta \rangle = a^{-1} \, \overline{\psi(a^{-1} r_{-\theta}(-\vec{b}))}. \tag{24}$$

• *Reproducing kernel:* taking as the signal the wavelet ψ itself, one obtains the reproducing kernel K, which measures the *correlation length* in each variable $\vec{b}, a, \theta$:

$$c_\psi \, K(\vec{b}, a, \theta | \vec{0}, 1, 0) = \langle \psi_{\vec{b},a,\theta} | \psi \rangle = a^{-1} \int \overline{\psi(a^{-1} r_{-\theta}(\vec{x} - \vec{b}))} \, \psi(\vec{x}) \, d^2\vec{x}. \tag{25}$$

• *Benchmark signals:* for testing particular properties of the wavelet, such as its ability to detect a discontinuity or its angular selectivity in detecting a particular direction, one may use appropriate 'benchmark' signals.

A typical example is the calibration of the angular selectivity of a directional wavelet. First one needs a suitable parameter. One that has proven efficient is the so-called *angle resolving power* (ARP) [4, 5]. This is defined as the opening angle of the cone tangent to the (effective) support of the wavelet in $\vec{k}$-space. For the Cauchy wavelet (22), the ARP is simply the opening angle $\Phi = 2\alpha$ of the supporting cone. Then one may use as benchmark a semi-infinite rod, sitting along the positive x-axis, and modeled with a delta function:

$$s(\vec{x}) = \vartheta(x) \, \delta(y), \tag{26}$$

where $\vartheta(x)$ is the step function.

Taking either a Morlet wavelet or a Cauchy wavelet, oriented at an angle θ, one computes the CWT of s as a function of x. The result is that both wavelets detect the orientation of the rod with a precision of the order of 5°. Indeed, for $\theta < 5°$, the WT is a "wall", increasing smoothly from 0, for $x \leq -5$, to its asymptotic value (normalized to 1) for $x \geq 5$. Then, for increasing misorientation θ, the wall gradually collapses, and essentially disappears for $\theta > 15°$. Only the tip of the rod remains visible, and for large θ ($\theta > 45°$), it gives a sharp peak (see Figure 2 [4, 5]).

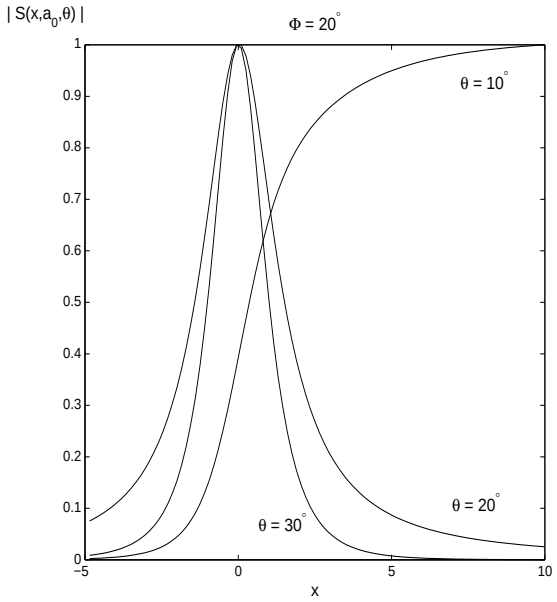

FIGURE 2. Testing the angular selectivity of the Cauchy wavelet $\widehat{\psi}_{44}$ of Figure 1 with the semi-infinite rod signal. The figure shows the modulus of the CWT as a function of $\vec{x}$, for various values of the misorientation angle θ.

5. GOING DISCRETE

5.1. Discretization of the CWT

The reproduction property (13) means that the information contained in the WT $S(\vec{b}, a, \theta)$ is highly redundant. This redundancy may be eliminated (this is the basic idea behind the *discrete* WT), or exploited, either under the form of interpolation formulas or for discretizing the reconstruction formula (11), as needed for numerical evaluation. The integral is replaced by a sum over a discrete (but infinite) family of wavelets $\psi_{\vec{b}_i, a_j, \theta_k}$, which can be chosen in such a way that no information is lost:

$$s(\vec{x}) = \sum_{ijk} \psi_{\vec{b}_i, a_j, \theta_k}(\vec{x})\ S(\vec{b}_i, a_j, \theta_k). \tag{27}$$

Such an overcomplete family is called a *frame*, according to the terminology introduced by Duffin and Schaefer [21] in the context of nonharmonic Fourier series. Its existence for specific wavelets may be proven along the same lines as in the 1-D case [17, 18, 19] with similar results [35]. In practical applications, the infinite sum will be truncated (a few terms will often suffice) and the approximate reconstruction so obtained is numerically stable [18, 19].

The problem, of course, is how to choose the sampling grid in an optimal fashion. The 2-D wavelet transform too obeys sampling theorems, that give

lower bounds on the density of sampling points, like the standard Shannon theorem of signal analysis, only more complicated. In practice, the sampling points are often fixed empirically, but a systematic exploitation of the reproducing kernel K leads to a minimal dicretization grid [4, 5].

5.2. The discrete WT in one dimension

One of the successes of the WT was the discovery that it is possible to construct functions ψ generating an orthonormal wavelet basis of $L^2(\mathbb{R})$. The construction is based on two facts: first, almost all examples of orthonormal bases of wavelets can be derived from a multiresolution analysis, and then the whole construction may be transcripted into the language of digital filters, familiar in the signal processing literature. As a consequence, it yields fast (pyramidal) algorithms, and this is the key to the usefulness of wavelets in many applications. In the 2-D case, the situation is exactly the same, as we shall sketch in this section. Further information may be found in [19] or [32].

In 1-D, a *multiresolution analysis* of $L^2(\mathbb{R})$ is an increasing sequence of closed subspaces

$$\ldots \subset V_{-2} \subset V_{-1} \subset V_0 \subset V_1 \subset V_2 \subset \ldots, \tag{28}$$

with $\bigcap_{j\in\mathbb{Z}} V_j = \{0\}$ and $\bigcup_{j\in\mathbb{Z}} V_j$ dense in $L^2(\mathbb{R})$, and such that

(1) $f(x) \in V_j \Leftrightarrow f(2x) \in V_{j+1}$

(2) There exists a real function $\phi \in V_0$, called a *scaling function*, such that the family $\{\phi(x-k), k \in \mathbb{Z}\}$ is an orthonormal basis of V_0.

Combining conditions (1) and (2), one gets an orthonormal basis of V_j, namely $\{\phi_{j,k}(x) \equiv 2^{j/2}\phi(2^j x - k), k \in \mathbb{Z}\}$.

Each V_j can be interpreted as an *approximation space*: the approximation of $f \in L^2(\mathbb{R})$ at the resolution 2^{-j} is defined by its projection onto V_j, and the larger j, the finer the resolution obtained. Then condition (1) means that no scale is privileged. The additional details needed for increasing the resolution from 2^{-j} to $2^{-(j+1)}$ are given by the projection of f onto the orthogonal complement W_j of V_j in V_{j+1}:

$$V_j \oplus W_j = V_{j+1}, \tag{29}$$

and we have:

$$L^2(\mathbb{R}) = \bigoplus_{j\in\mathbb{Z}} W_j = V_{j_o} \oplus \left(\bigoplus_{j=j_o}^{\infty} W_j \right), \tag{30}$$

where j_o is an arbitrary lowest resolution level. Then the theory asserts the existence of a function ψ, called the *mother wavelet*, explicitly computable from ϕ, such that $\{\psi_{j,k}(x) \equiv 2^{j/2}\psi(2^j x - k),\ j,k \in \mathbb{Z}\}$ constitutes an orthonormal basis of $L^2(\mathbb{R})$: these are the *orthonormal wavelets*.

The construction of ψ proceeds as follows. First, the inclusion $V_0 \subset V_1$ yields the relation (called the scaling or refining equation):

$$\phi(x) = \sqrt{2} \sum_{n=-\infty}^{\infty} h_n \phi(2x - n), \quad h_n = \langle \phi_{1,n} | \phi \rangle. \tag{31}$$

Taking Fourier transforms, this gives

$$\widehat{\phi}(2\omega) = m_0(\omega)\, \widehat{\phi}(\omega), \quad \text{with} \quad m_0(\omega) = \frac{1}{\sqrt{2}} \sum_{n=-\infty}^{\infty} h_n e^{-in\omega}. \tag{32}$$

Thus m_0 is a 2π-periodic function and it satisfies the relation

$$|m_0(\omega)|^2 + |m_0(\omega + \pi)|^2 = 1, \quad \text{a.e.}. \tag{33}$$

Iterating (32), one gets the scaling function as the (convergent!) infinite product

$$\widehat{\phi}(\omega) = (2\pi)^{-1/2} \prod_{j=1}^{\infty} m_0(2^{-j}\omega). \tag{34}$$

Then one defines the function $\psi \in W_0 \subset V_1$ by the relation

$$\widehat{\psi}(2\omega) = m_1(\omega)\, \widehat{\phi}(\omega), \tag{35}$$

where m_1 is another 2π-periodic function. By the relation (29) and the orthonormality of the functions $\{\phi_{j,k}\}$, the functions m_0, m_1 must satisfy the identity

$$m_1(\omega)\, \overline{m_0(\omega)} + m_1(\omega + \pi)\, \overline{m_0(\omega + \pi)} = 0, \quad \text{a.e.}. \tag{36}$$

The simplest solution is to put $m_1(\omega) = e^{i\omega}\ \overline{m_0(\omega + \pi)}$, which implies, in particular $|m_0(\omega)|^2 + |m_1(\omega)|^2 = 1$, a.e.. Then one obtains

$$\psi(x) = \sqrt{2} \sum_{n=-\infty}^{\infty} (-1)^{n-1} h_{-n-1} \phi(2x - n), \tag{37}$$

and one proves that this function indeed generates an orthonormal basis with all the required properties. Various additional conditions may then be imposed on the basic wavelet ψ: arbitrary regularity, several vanishing moments (in any case, ψ has always mean zero), symmetry, fast decrease at infinity, and even compact support [19].

Actually, the discussion above means that we have translated the multiresolution structure into the language of digital filters (by a filter, we mean a multiplication operator in frequency space or a linear convolution in the time variable). For instance, $m_0(\omega)$ is a filter, with Fourier coefficients h_n, $m_1(\omega)$ is another one, and $\{m_0, m_1\}$ are called Quadrature Mirror Filters or QMF. Then the various restrictions imposed on ψ translate into suitable constraints on the filter coefficients h_n. For instance, ψ has compact support if only finitely many

h_n differ from zero (one then speaks of a finite impulse response or FIR filter). The rapidity of the algorithms depends crucially on the length of the filters involved, because the pyramidal structure rests on a concatenation of several filters.

However, it turns out that the scheme based on orthonormal wavelet bases is too rigid for most applications and various generalizations have been proposed. Among them, we may quote: biorthogonal wavelet bases [14], wavelet packets and the best basis algorithm [15, 32, 40], second generation wavelets and the so-called 'lifting scheme' [37], integer wavelet transforms [13].

5.3. The discrete WT in two dimensions

In 2-D, the simplest approach consists in building a multiresolution analysis simply by taking the direct (tensor) product of two such structures in 1-D, one for the x direction, one for the y direction. If $\{V_j, j \in \mathbb{Z}\}$ is a multiresolution analysis of $L^2(\mathbb{R})$, then $\{\widetilde{V}_j = V_j \otimes V_j, j \in \mathbb{Z}\}$ is a multiresolution analysis of $L^2(\mathbb{R}^2)$. Writing again $\widetilde{V}_j \oplus \widetilde{W}_j = \widetilde{V}_{j+1}$, it is easy to see that this 2-D analysis requires one scaling function : $\Phi(x,y) = \phi(x)\,\phi(y)$, but three wavelets:

$$\Psi^h(x,y) = \phi(x)\,\psi(y),\ \Psi^v(x,y) = \psi(x)\,\phi(y),\ \Psi^d(x,y) = \psi(x)\,\psi(y). \qquad (38)$$

As the notation suggests, Ψ^h detects preferentially horizontal edges, that is, discontinuities in the vertical direction, whereas Ψ^v and Ψ^d detect vertical and oblique edges, respectively.

From these three wavelets, one gets an orthonormal basis of $\widetilde{V}_j$ by defining $\{\Phi^j_{kl}(x,y) = \phi_{j,k}(x)\,\phi_{j,l}(y), k,l \in \mathbb{Z}\}$, and one for $\widetilde{W}_j$ in the same way, namely $\{\Psi^{\alpha,j}_{kl}(x,y), \alpha = h,v,d \text{ and } k,l \in \mathbb{Z}\}$. Clearly this construction enforces a Cartesian geometry, with the horizontal and the vertical directions playing a preferential role. This is natural for certain types of images, such as in television, but is poorly adapted for detecting edges in arbitrary directions.

As in the 1-D case, the implementation of this construction rests on a pyramidal algorithm introduced by Mallat [28, 29]. The technique consists again in translating the multiresolution structure into the language of QMF, and putting suitable constraints on the filter coefficients.

5.4. Fast algorithms for the CWT

Besides the full discretization described in Section 5.1, and the discrete WT just discussed, there is an intermediate procedure, introduced in [22], under the name of infinitesimal multiresolution analysis. It consists in discretizing the scale variable alone, on an arbitrary sequence of values (not necessarily powers of a fixed ratio). This leads to fast algorithms that could put the CWT on the same footing as the DWT in terms of speed and efficiency, by extending the advantages of the latter to cases where no exact QMF is available. Let us sketch the method, first in 1-D. Further details may be found in [38].

Instead of the standard L^2-normalization used in (1), it is more convenient to choose the L^1-normalization, namely to use $\psi_{(b,a)} = a^{-1}\psi\left(a^{-1}(t-b)\right)$. Then, given a wavelet ψ, normalized to $c_\psi = 1$, one lumps together all the low frequency components in a scaling function

$$\Phi(t) = \int_1^\infty \psi\left(\frac{t}{a}\right)\frac{da}{a^2} = \frac{1}{t}\int_0^t \psi(s)\,ds, \qquad \widehat{\Phi}(\omega) = \int_1^\infty \widehat{\psi}(a\omega)\,\frac{da}{a}, \tag{39}$$

and introduces the integrated wavelet

$$\Psi(t) = \int_{1/2}^1 \psi\left(\frac{t}{a}\right)\frac{da}{a^2} = \frac{1}{t}\int_t^{2t} \psi(s)\,ds, \qquad \widehat{\Psi}(\omega) = \int_{1/2}^1 \widehat{\psi}(a\omega)\,\frac{da}{a}. \tag{40}$$

These functions satisfy two-scale relations:

$$\Psi(t) = 2\Phi(2t) - \Phi(t), \quad \widehat{\Psi}(\omega) = \widehat{\Phi}(\omega/2) - \widehat{\Phi}(\omega). \tag{41}$$

Next, one chooses a regular grid, as opposed to the dyadic one used in the discrete case, namely:

$$\Phi_t^j \equiv \Phi_{(t,2^{-j})} = 2^j\Phi(2^j(\cdot - t)), \quad \Psi_t^j = 2^j\Psi(2^j(\cdot - t)). \tag{42}$$

Although the resulting transform will be redundant, it has the great advantage over the conventional DWT of maintaining translation covariance. Then, exactly as in (30), one gets a discrete reconstruction formula:

$$s(t) = \langle \Phi_t^{j_o} | s\rangle + \sum_{j=j_o}^{\infty} \langle \Psi_t^j | s\rangle. \tag{43}$$

Then assume there exists two functions μ_0, μ_1 satisfying the following relations, analogous to (32), (35),

$$\widehat{\Phi}(2\omega) = \mu_0(\omega)\widehat{\Phi}(\omega), \quad \widehat{\Psi}(2\omega) = \mu_1(\omega)\widehat{\Phi}(\omega), \quad \text{a.e.} \tag{44}$$

These functions are *not* necessarily 2π-periodic. However, since using the regular grid means sampling $\Phi(t)$ at unit rate, we have to assume that the function $\widehat{\Phi}$ is essentially supported in $[-\pi, \pi]$. Therefore, in view of the relations (44), it is reasonable to approximate the functions μ_0, μ_1 by 2π-periodic functions m_0, m_1. In fact in can be shown [38] that there exists a unique pair m_0, m_1 that minimizes the quantities

$$\nu(\mu_i, m_i) = \left[\int_{\mathbb{R}} |(\mu_i(\omega) - m_i(\omega))\widehat{\Phi}(\omega)|^2\right]^{1/2}, \quad i = 0, 1,$$

namely

$$m_0(\omega) \;=\; \frac{\sum_{k\in\mathbb{Z}} \overline{\widehat{\Phi}(\omega + 2k\pi)}\,\widehat{\Phi}(2\omega + 4k\pi)}{\sum_{k\in\mathbb{Z}} |\widehat{\Phi}(\omega + 2k\pi)|^2}, \tag{45}$$

$$m_1(\omega) \quad = \quad \frac{\sum_{k\in\mathbb{Z}} \overline{\widehat{\Phi}(\omega+2k\pi)}\,\widehat{\Psi}(2\omega+4k\pi)}{\sum_{k\in\mathbb{Z}} |\widehat{\Phi}(\omega+2k\pi)|^2}. \tag{46}$$

These approximate filters m_0, m_1, which are called *pseudo-QMF*, satisfy the identity $m_0(\omega) + m_1(\omega) = 1$.

More flexibility is obtained if one subdivides the scale interval $[1/2, 1]$ into n subbands, by $a_o = 1/2 < a_1 < \ldots < a_n = 1$. In that case one ends up with one scaling function $\Phi(t)$ and n integrated wavelets $\Psi_i(t)$, $i = 0, \ldots n-1$, corresponding to integration from a_{i-1} to a_i. An additional improvement consists in periodizing the signal and computing filters m_0, m_1 of the same length as the signal. The resulting pyramidal algorithm has a complexity equal to one half of the traditional FFT value $\mathcal{O}(N \log_2^2 N)$. Thus one obtains a very fast implementation of the CWT, truly competitive with the DWT.

The extension to 2-D is straightforward. Starting from an isotropic wavelet ψ, one gets a scaling function $\widehat{\Phi}(\vec{k})$ and a family of isotropic integrated wavelets $\widehat{\Psi}_i^{\rm iso}(\vec{k}) \equiv \widehat{\Psi}_i^{\rm iso}(k)$, where $k = |\vec{k}|$. But one can do better and design directional pseudo-QMF as follows. Let $\{\eta_l(\theta), l = 1, \ldots d\}$ be a resolution of the identity consisting of C^∞, 2π-periodic, functions of compact support, i.e. $\sum_{l=1}^d \eta_l(\theta) = 1$. Then one obtains a family of directional integrated wavelets, in polar coordinates, as

$$\widehat{\Psi}_{i,l}(k,\theta) = \widehat{\Psi}_i^{\rm iso}(k)\,\eta_l(\theta), \tag{47}$$

and indeed one has

$$\sum_{l=1}^{d} \widehat{\Psi}_{i,l}(k,\theta) = \widehat{\Psi}_i^{\rm iso}(k). \tag{48}$$

The net result is a discrete, fast, implementation of the 2-D CWT, including the directional degree of freedom (here the last improvement described above in the 1-D case becomes crucial). The preliminary applications of this algorithm look very promising, for instance in directional filtering [24, 39].

6. The group-theoretical background: Continuous wavelets as affine coherent states

As already mentioned in Section 2, the CWT is entirely rooted in group theory. Indeed, the operations of translations, rotations and global dilations make up the so-called *similitude* group of the plane (or Euclidean group with dilations), $SIM(2) = \mathbb{R}^2 \rtimes (\mathbb{R}_*^+ \times SO(2))$ ($\rtimes$ denotes a semidirect product). Then the relation

$$(U(\vec{b}, a, \theta)s)(\vec{x}) = s_{\vec{b},a,\theta}(\vec{x}) = a^{-1}\, s(a^{-1}\, r_{-\theta}(\vec{x} - \vec{b})), \tag{49}$$

defines the natural representation of $SIM(2)$ in the Hilbert space $L^2(\mathbb{R}^2, d^2\vec{x})$ of finite energy signals, and it is unitary and irreducible. Furthermore, U is also square integrable, which means there exists at least one (and in fact a dense set of) admissible vectors, i.e. vectors ψ such that the matrix element

$\langle U(\vec{b}, a, \theta)\psi|\psi\rangle$ is square integrable over the group, with respect to the Haar measure $dg \equiv a^{-3} d^2\vec{b}\, da\, d\theta$). Indeed a straightforward calculation shows that

$$\iiint |\langle U(a, \theta, \vec{b})\psi|\psi\rangle|^2 \, a^{-3} d^2\vec{b}\, da\, d\theta = c_\psi \|\psi\|^2, \tag{50}$$

where c_ψ is the constant defined in (5), so that the two notions of admissibility indeed coincide. From this, one can derive all the properties of the 2-D CWT described in Section 2 [1, 2].

Now this is not an isolated fact, but an example of a general pattern. Indeed, let $\mathcal{H} \equiv L^2(Y, d\mu)$ be the space of finite energy signals on a manifold Y, and assume there is a transformation group G acting on Y, with a continuous unitary irreducible representation U in $\mathcal{H}$. Assume furthermore that the representation U is square integrable, that is, there exists at least one nonzero admissible vector $\psi \in \mathcal{H}$. Under these conditions, a *G-adapted wavelet analysis* on Y may be constructed, following the general construction of coherent states associated to G, that we now sketch (see [1, 2] for details).

Choose a fixed admissible vector $\psi \in \mathcal{H}$ (the analyzing wavelet). Then the wavelets are the vectors $\psi_g = U(g)\psi \in \mathcal{H}$ $(g \in G)$, and the corresponding continuous wavelet transform (CWT) is defined as:

$$S_\psi(g) = \langle \psi_g | s \rangle \tag{51}$$

Introduce again the linear map $W_\psi : \mathcal{H} \to L^2(G, dg)$ given by $(W_\psi s)(g) \equiv c_\psi^{-1/2} S_\psi(g)$, where

$$c_\psi = \int_G |\langle U(g)\psi|\psi\rangle|^2 \, dg \tag{52}$$

and dg denotes the left invariant Haar measure on G. Then the CWT has the following properties [1, 2], that match exactly those described in Section 2.1:

(i) The CWT is *covariant* under the action of the group G:

$$W_\psi[U(g)s](g_o) = (W_\psi s)(g^{-1} g_o), \ \forall\, g \in G. \tag{53}$$

(ii) *Norm conservation:*

$$c_\psi^{-1} \int_G |S_\psi(g)|^2 \, dg = \int_Y |s(y)|^2 \, d\mu(y), \tag{54}$$

i.e. W_ψ is an isometry; hence its range, the space of wavelet transforms, is a *closed* subspace $\mathcal{H}_\psi$ of $L^2(G, dg)$.

(iii) By (i), W_ψ may be inverted on its range by the transposed map, which gives the *reconstruction formula:*

$$s(y) = c_\psi^{-1} \int_G S_\psi(g)\psi_g(y) \, dg. \tag{55}$$

(iv) The projection from $L^2(G, dg)$ onto $\mathcal{H}_\psi$ is an integral operator with kernel $K(g, g') = c_\psi^{-1}\langle\psi_g|\psi_{g'}\rangle$, that is, the auto-correlation function of ψ, also called a *reproducing kernel;* in other words, a function $f \in L^2(G, dg)$ is a WT iff it satisfies the reproducing relation:

$$f(g) = c_\psi^{-1} \int_G \langle\psi_g|\psi_{g'}\rangle f(g')\, dg'. \tag{56}$$

Now it may happen that the analyzing wavelet ψ has a nontrivial isotropy subgroup H_ψ, up to a phase, i.e.

$$U(h)\psi = e^{i\alpha(h)}\psi,\ h \in H_\psi. \tag{57}$$

In this case, the whole construction may be performed [1, 2] under a slightly less restrictive condition, namely the representation U need only to be square integrable on the coset space $X = G/H_\psi$. Then one obtains wavelets indexed by the points of X, namely $\psi_x = U(\sigma(x))\psi\ (x \in X)$, where $\sigma : X \to G$ is an *arbitrary* section.

The interested reader may find the detailed theory in the review [1] and papers quoted there. It is then easy to see that all the properties of the 2-D CWT discussed in Section 2 are simply the particularization to the group $SIM(2)$ of those listed above.

The advantage of this point of view is twofold. First, it shows that the CWT is firmly based on mathematical theorems, not numerical recipes. Then, it allows the extension of the CWT to a host of other situations, such as 3 space dimensions, the 2-sphere and similar manifolds, and also to space-time (time-dependent signals or images, such as TV or video sequences), including relativistic effects (using wavelets associated to the affine Galilei or Poincaré group). We refer the reader to [10, Chapter 2] or to [1, 2] for further details.

7. Applications of the 2-D CWT

The 2-D CWT has been used by a number of authors, in a wide variety of physical problems [16, 33, 34]. In all cases, its main use is for the *analysis* of images. It can be used for the detection of specific features, such as a hierarchical structure, edges, filaments, contours, boundaries between areas of different luminosity, etc. Of course, the type of wavelet chosen depends on the precise aim. An isotropic wavelet (mexican hat) suffices for pointwise analysis, but an oriented wavelet (Morlet, Cauchy) is more efficient for the detection of oriented features in the signal, that is, regions where the amplitude is regular along one direction and has a sharp variation along the perpendicular direction.

In this section, we will quickly list the most significant applications. We refer the reader to [10], in particular Chapter 2, for a detailed survey, including the original references.

FIGURE 3. CWT of a thick letter 'A', with a mexican hat and $a = 0.075$, in level curves. (a) the letter; (b) the WT, showing the contours.

7.1. Pointwise analysis

- Contour detection, character recognition: since the CWT is sensitive to discontinuities, it is very efficient for detecting the *contour* [3, 35] or the *edges* of an object [25, 31] (which are discontinuities in luminosity). An immediate application is automatic character recognition in optical reading (see Figure 3).
- Analysis of 2-D fractals [8], either artificial (numerical snowflakes, diffusion limited aggregates) or natural (electrodeposition clusters, various arborescent phenomena); here the (a, θ) representation is useful, since it presents the signal at all scales at once; particular applications include the measurement of the fractal dimensions and the unraveling of universal laws (mean angle between branches, azimuthal Cantor structures, ...).
- Analysis of astronomical images: the CWT has been used for several purposes, such as removal of background sky, unraveling of the hierarchical structure of a galactic nebula, or that of the universe itself (galaxy counts, detection of galaxy clusters or voids); detection of Einstein gravitational arcs in cosmological pictures, with an annular wavelet.
- Determination of the local regularity of a signal, by estimation of local Lipschitz exponents.
- Medical physics and psychophysics: modelling of human vision, e.g. definition of local contrast in images (actually a nonlinear extension of the CWT), medical imaging, in particular 2-D NMR imaging and tomography.

7.2. Applications of directional wavelets

- Fluid dynamics: the CWT has been applied succesfully to the analysis of 2-D developed turbulence in fluids, in particular the localization of small scales in the distribution of energy or enstrophy [23]; other applications in

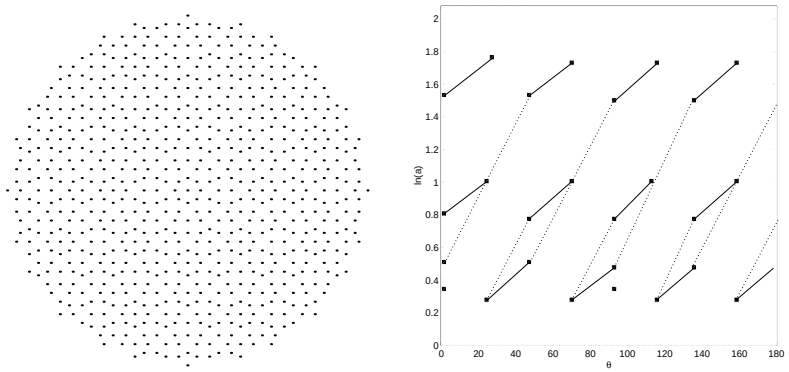

FIGURE 4. Analysis of an octagonal pattern : (a) the pattern; (b) the local maxima of its scale angle measure $\mu_S(a,\theta)$; this pattern has a rotation symmetry by $\pi/4$, and two distinct mixed symmetries, consisting of a rotation by $\pi/8$ combined with a dilation by $\delta_1 = \sqrt{2}\cos(\pi/8)$, resp. $\delta_2 = 2\cos(\pi/8)$. Homologous maxima are linked by a line segment, continuous for δ_1 and dashed for δ_2.

fluid dynamics include the visualization and measurement of a velocity field with help of an directional wavelet, or the disentangling of a wave train.

– Detection of symmetries: directional wavelets may be used for detecting (hidden) dilation-rotation symmetries in patterns, such as Penrose tilings or the diffraction spectrum of a quasi-crystal [6, 39]. The tool here is the so-called *scale-angle measure* of the signal, namely the positive function

$$\mu_S(a,\theta) = \int |S(\vec{b},a,\theta)|^2 \, d^2\vec{b} \;\sim\; \int |\widehat{\psi}(ar_{-\theta}(\vec{k}))|^2 \, |\widehat{s}(\vec{k})|^2 \, d^2\vec{k}. \qquad (58)$$

Figure 4 shows an example of this analysis.

– Other applications of the CWT with directional wavelets include the analysis of geological faults, or that of textures, both of which usually carry oriented features.

By contrast, the DWT is used mainly whenever data compression is essential, in particular for the reconstruction of a signal after some kind of preprocessing. Typical applications would be, for instance, the representation of images in terms of wavelet maxima, image compression and coding (e.g. in HDTV), image and signal denoising, or identification of fingerprints (the algorithm used by the FBI [11]).

8. CONCLUSION

As in 1-D signal analysis, wavelet techniques have become an established tool in image processing, both in their DWT and CWT incarnations and their generalizations. Actually the DWT and the CWT have almost opposite properties,

hence their ranges of application differ widely too. The DWT and its generalizations are fast and economical, they yield for instance impressive data compression rates, which is especially useful in image processing, where huge amounts of data, mostly redundant, have to be stored and transmitted. It is also the most popular wavelet technique. On the other hand, the CWT is very efficient at detecting specific features in signals or images, such as in pattern recognition or directional filtering, and thus it is often a better tool for analysis. Furthermore, fast CWT algorithms are increasingly available, and this should dispel the wrong, but widely held, belief that the CWT is too cost intensive for practical applications.

References

1. S.T. Ali, J-P. Antoine, J-P. Gazeau and U.A. Mueller (1995). Coherent states and their generalizations: A mathematical overview, *Reviews Math. Phys.* **7**, 1013–1104.
2. S.T. Ali, J-P. Antoine and J-P. Gazeau (1999). *Coherent States, Wavelets and their Generalizations.* Springer, New York and Berlin (to appear).
3. J-P. Antoine, P. Carrette, R. Murenzi and B. Piette (1993). Image analysis with two-dimensional continuous wavelet transform, *Signal Proc.* **31**, 241–272.
4. J-P. Antoine and R. Murenzi (1995). Two-dimensional directional wavelets and the scale-angle representation, *Signal Proc.* **52**, 259–281.
5. J-P. Antoine, R. Murenzi and P. Vandergheynst (1996). Two-dimensional directional wavelets in image processing, *Int. J. of Imaging Systems and Technology.* **7**, 152–165.
6. J-P. Antoine, R. Murenzi and P. Vandergheynst (1998). Directional wavelets revisited: Cauchy wavelets and symmetry detection in patterns, *Applied Comput. Harm. Anal.* (to appear).
7. J-P. Antoine and P. Vandergheynst (1998). Wavelets on the 2-sphere: A group-theoretical approach, *Applied Comput. Harm. Anal.* (to appear).
8. A. Arnéodo, F. Argoul, E. Bacry, J. Elezgaray, E. Freysz, G. Grasseau, J.F. Muzy and B. Pouligny (1991). Wavelet transform of fractals, in [33], pp. 286–352.
9. A. Arnéodo, E. Bacry, P.V. Graves and J.F. Muzy (1996). Characterizing long-range correlations in DNA sequences from wavelet analysis, *Phys. Rev. Lett.* **74**, 3293–3296.
10. J.C. van den Berg (ed.) (1998). *Wavelets in Physics.* Cambridge Univ. Press, Cambridge.
11. C.M. Brislawn (1995). Fingerprints go digital, *Notices Amer. Math. Soc.* **42**, 1278–1283.
12. B. Burke Hubbard (1998). *The World According to Wavelets* (2nd ed.). A.K. Peters, Wellesley, MA.
13. A.R. Calderbank, I. Daubechies, W. Sweldens and B.L. Yeo

(1998). Wavelets that map integers to integers, *Applied Comput. Harm. Anal.* **5**, 332–369.
14. A. Cohen, I. Daubechies and J-C. Feauveau (1992). Biorthogonal bases of compactly supported wavelets, *Comm. Pure Appl. Math.* **45**, 485–560.
15. R.R. Coifman, Y. Meyer, S. Quake and M.V. Wickerhauser (1993). Signal processing and compression with wavelet packets, in [34], pp. 77–93.
16. J-M. Combes, A. Grossmann and Ph. Tchamitchian (eds.) (1990). *Wavelets, Time-Frequency Methods and Phase Space (Proc. Marseille 1987), 2d ed.* Springer, Berlin.
17. I. Daubechies, A. Grossmann and Y. Meyer (1986). Painless nonorthogonal expansions, *J. Math. Phys.* **27**, 1271–1283.
18. I. Daubechies (1990). The wavelet transform, time-frequency localization and signal analysis, *IEEE Trans. Inform. Theory.* **36**, 961–1005.
19. I. Daubechies (1992). *Ten Lectures on Wavelets.* SIAM, Philadelphia, PA.
20. R. De Valois and K. De Valois (1988). *Spatial Vision.* Oxford Univ. Press, New York.
21. R.J. Duffin and A.C. Schaefer (1952). A class of nonharmonic Fourier series, *Trans. Amer. Math. Soc.* **72**, 341–366.
22. M. Duval-Destin, M-A. Muschietti and B. Torrésani (1993). Continuous wavelet decompositions, multiresolution, and contrast analysis, *SIAM J. Math. Anal.* **24**, 739–755.
23. M. Farge (1992). Wavelet transforms and their applications to turbulence, *Annu. Rev. Fluid Mech.* **24**, 395–457.
24. J-F. Gobbers and P. Vandergheynst (1998). *A fast continuous wavelet transform*, preprint UCL-IPT-98-07, Louvain-la-Neuve (in preparation).
25. A. Grossmann (1988). Wavelet transform and edge detection, in *Stochastic Processes in Physics and Engineering,* pp.149–157; S. Albeverio, Ph. Blanchard, L. Streit and M. Hazewinkel (eds.). Reidel, Dordrecht.
26. M. Holschneider (1991). Inverse Radon transforms through inverse wavelet transforms, *Inverse Problems.* **7**, 853–861.
27. G. Kaiser and R.F. Streater (1992). Windowed Radon transforms, analytic signals, and the wave equation, in *Wavelets: A Tutorial in Theory and Applications,* pp.399–441; C. K. Chui (ed.). Academic Press, London.
28. S.G. Mallat (1989). Multifrequency channel decompositions of images and wavelet models, *IEEE Trans. Acoust., Speech, Signal Proc.* **37**, 2091–2110.
29. S.G. Mallat (1989). A theory for multiresolution signal decomposition: the wavelet representation, *IEEE Trans. Pattern Anal. Machine Intell.* **11**, 674–693.
30. D. Marr (1982). *Vision.* Freeman, San Francisco.
31. D. Marr and E. Hildreth (1980). Theory of edge detection, *Proc. R. Soc. Lond. B.* **207**, 187–217.
32. Y. Meyer (1993). *Wavelets, Algorithms and Applications.* SIAM, Philadelphia, PA.

33. Y. MEYER (ed.) (1991). *Wavelets and Applications (Proc. Marseille 1989).* Springer, Berlin, and Masson, Paris.
34. Y. MEYER and S. ROQUES (eds.) (1993). *Progress in Wavelet Analysis and Applications (Proc. Toulouse 1992).* Ed. Frontières, Gif-sur-Yvette.
35. R. MURENZI (1990). Ondelettes multidimensionnelles et applications à l'analyse d'images, Thèse de Doctorat, Univ. Cath. Louvain.
36. E.M. STEIN and G. WEISS (1971). *Introduction to Fourier Analysis on Euclidean Spaces.* Princeton Univ. Press, Princeton, NJ.
37. W. SWELDENS (1996). The lifting scheme: a custom-design construction of biorthogonal wavelets, *Applied Comput. Harm. Anal.* **3**, 1186–1200.
38. B. TORRÉSANI (1995). *Analyse continue par ondelettes.* InterÉditions / CNRS Éditions, Paris.
39. P. VANDERGHEYNST (1998). Ondelettes directionnelles et ondelettes sur la sphère, Thèse de Doctorat, Univ. Cath. Louvain.
40. M.V. WICKERHAUSER (1994). *Adapted Wavelet Analysis from Theory to Software.* A.K. Peters, Wellesley, MA.
41. R.A. ZUIDWIJK (1997). The discrete and continuous wavelet x-ray transform, *SPIE Proc. 3169, Wavelet Applications in Signal and Image Processing* **V**: 357–366.