Two Algorithms for Non-Separable Wavelet Transforms and Applications to Image Compression

Franklin Mendivil* and Daniel Piché

Department of Applied Mathematics
University of Waterloo
Waterloo, Ontario, Canada N2L 3G1

Summary. Previously, the use of non-separable wavelets in image processing has been hindered by the lack of a fast algorithm to perform a non-separable wavelet transform. We present two such algorithms in this paper. The first algorithm implements a periodic wavelet transform for any valid wavelet filter sequence and dilation matrices satisfying a trace condition. We discuss some of the complicating issues unique to the non-separable case and how to overcome them. The second algorithm links Haar wavelets and complex bases and uses this link to drive the algorithm. For the complex bases case, the asymmetry of the wavelet trees produced leads to a discussion of the complexities in implementing zero-tree and other wavelet compression methods. We describe some preliminary attempts at using this algorithm, with non-separable Haar wavelets, for reducing the blocking artifacts in fractal image compression.

Keywords: Haar, wavelet, image, compression, fractal, tiling, periodic wavelet transform, non-separable.

1 Introduction

Wavelets have found applications to many diverse areas of science, engineering and mathematics. A large part of their usefulness in applications is their ability to capture both scale and location information of a data signal. However, the practical application of wavelets is also driven by the existence of a fast algorithm to do a wavelet decomposition on a discrete data set. This *Fast Wavelet Transform* in many cases allows the real-time analysis of data.

For a finite length data signal, we have several options on how to analyze the data. One popular method is to periodize the signal. In this case, we think of our data as being a function on the circle S^1 and use wavelets on S^1 that are induced from wavelets on $\mathbb{R}$. There is a corresponding discrete transform called the *Discrete Periodic Wavelet Transform* that is very fast. In fact, for a data stream of length N, the transform has time complexity $O(N)$ see [30]). Another method is *zero padding* where you extend your finite length data signal by zeros to make an (potentially) infinite length signal. In processing zero padded data you

* Current address: School of Math, Georgia Tech, Atlanta, GA 30332-0160

need only compute with the finite length signal itself, filling in the appropriate zeros where necessary.

One of the many areas in which wavelets have found applications is the area of image processing. Here their ability to extract spatially localized information is very useful in reconstructing, modifying or analyzing an image. However, most of the wavelets used in image processing (an inherently two-dimensional application) have been tensor products of wavelets from $L^2(\mathbb{R})$. While this is sufficient for many purposes, in some cases it is not. For example, when using tensor product wavelets to compress images, you often introduce artifacts from the fact that you lose information. With tensor product wavelets, these artifacts include vertical and horizontal lines in the image. This is undesirable since our eyes are particularly sensitive to errors along lines. Non-separable wavelets based on fractal or dust-like tilings introduce a natural dithering effect which helps to eliminate these linear errors.

Non-separable wavelets are wavelets that are, in some sense, intrinsic to two (or more) dimensions; they are not tensor products of wavelets on some lower dimensional space. There has been much recent activity on constructing and analyzing multidimensional non-separable wavelets (see [2, 5, 19–21]). However, multidimensional non-separable wavelets are far from being well-understood. The question of existence and properties of these wavelets is a much more delicate and intricate one than that of one dimensional wavelets. For instance, there are very few explicit constructions of these wavelets. On the other hand, certain types of wavelets, namely Haar wavelets, are fairly easy to construct. These wavelets are derived from characteristic functions.

The characterization of multidimensional Haar wavelets was given in the paper of Gröchenig and Madych [15]. These wavelets are usually non-separable and have support on fractal tilings. This leads to the idea of complex bases, which also produce fractal tilings of the complex plane.

Complex bases are a way of representing complex numbers, in a similar fashion as the decimal system is used to represent real numbers. The study of such bases began with the work of Kátai and Szabó [18]. Many results in this area, including algorithms for determining the representations, are due to Gilbert [9–11]. Gilbert also provided the connection between the fractal tiles of complex bases and iterated function systems. This allowed the development of the long division algorithm for complex bases [14].

In this paper we present two recent algorithms to perform non-separable wavelet transforms, one a periodic transform and one a zero padding transform. The periodic transform requires that the dilation matrix satisfy a trace condition but will work with any valid wavelet filter sequence for the given dilation matrix. Thus it will work for wavelets more general than Haar wavelets. The zero padding algorithm is based on a new link between complex bases and wavelets, which enables an understanding of the translation of the Mallat algorithm to the language of complex bases. This algorithm only works for Haar wavelets. However, the dilation matrices associated with complex bases do not in general satisfy the trace condition. Thus, while the two algorithms have a substantial

area of overlap, they complement each other in that neither one generalizes the other.

In some sense, our periodic transform is simply the transform on the n dimensional torus induced by the discrete wavelet transform on $\mathbb{R}^n$ and periodicity. Unlike the one dimensional case, however, this turns out to have surprising complications so that it doesn't work out as simply. We indicate in this paper how these difficulties may be overcome.

After presenting each algorithm, we give some preliminary experiments in using non-separable Haar wavelets combined with a fractal-wavelet transform in an attempt to reduce the blocking artifacts which are common in conventional fractal image compression. In the case of the periodic transform, we use the Haar wavelets primarily because of a lack of suitable wavelet filters for other smoother wavelets. However, the benefits of non-separable wavelets are evident even in the simple Haar wavelet case.

2 Wavelets

In this section, we summarize the necessary background of wavelet analysis. For a more complete account of the wavelet theory the reader is referred to [6, 15] and of the fractal compression theory see [8, 22, 29].

To start, let $L^2(\mathbb{R}^n)$ be the space of all square Lebesgue integrable functions from $\mathbb{R}^n$ to $\mathbb{R}$. Recall the following definitions:

Definition 21 A matrix A on $\mathbb{R}^n$ is an *acceptable dilation for* $\mathbb{Z}^n$ if $A\mathbb{Z}^n \subset \mathbb{Z}^n$ and if $|\lambda| > 1$ for each eigenvalue λ of A.

Throughout this paper we will let $q = |\det A|$. The properties of an acceptable dilation imply that q is an integer ≥ 2.

Let A be an acceptable dilation on $L^2(\mathbb{R}^n)$, $f \in L^2(\mathbb{R}^n)$ and $x, y \in \mathbb{R}^n$. Define the unitary dilation operator U_A by

$$U_A f(x) = |\det A|^{-1/2} f(A^{-1}x)$$

and the translation operator τ_y by

$$\tau_y f(x) = f(x - y).$$

Then for each $i \in \mathbb{Z}$ and $j \in \mathbb{Z}^n$ let $f_{i,j} \equiv U_A^{-i} \tau_j f$. Hence,

$$f_{i,j}(x) = q^{i/2} f(A^i x - j).$$

Using this notation, $f_{0,0} = f$.

We are interested in wavelet bases given by translation by integers. By a basis we will mean a (orthonormal) Hilbert space basis.

Definition 22 A *wavelet basis* B, associated with an acceptable dilation A, is a basis of $L^2(\mathbb{R}^n)$ whose members are A dilates and $\mathbb{Z}^n$ translates of a finite orthonormal set $S = \{\psi^1, \ldots, \psi^m\} \subset L^2(\mathbb{R}^n)$, where $m \in \mathbb{N}^+$. More precisely,

$$B = \{\psi^l_{i,j} \ : \ l = 1, \ldots, m; \ i \in \mathbb{Z}; \ j \in \mathbb{Z}^n\},$$

where $\psi^l_{i,j} = q^{i/2}\psi^l(A^i x - j)$. The elements of S are called the *basic (mother) wavelets.*

Consider the definition of a multiresolution analysis as given in [15] where the lattice $\Gamma = \mathbb{Z}^n$.

Definition 23 Let A be an acceptable dilation for $\mathbb{Z}^n$. A *multiresolution analysis (MRA)* associated with A is a sequence of closed subspaces $(V_i)_{i\in\mathbb{Z}}$ of $L^2(\mathbb{R}^n)$, satisfying

i) $V_i \subset V_{i+1}$, $\forall i \in \mathbb{Z}$
ii) $\cup_{i\in\mathbb{Z}} V_i = L^2(\mathbb{R}^n)$
iii) $V_i = U_A^{-i} V_0$, $\forall i \in \mathbb{Z}$
iv) $\tau_j V_0 = V_0$, $\forall j \in \mathbb{Z}^n$
v) there is a function $\phi \in V_0$, called the *scaling function*, such that $\{\tau_j \phi \ : \ j \in \mathbb{Z}^n\}$ is a basis for V_0.

These properties imply that $\{\phi_{i,j} : j \in \mathbb{Z}^n\}$ is a basis for V_i for each $i \in \mathbb{Z}$. Since $\phi \in V_0 \subset V_1$, we obtain the *dilation equation*:

$$\begin{aligned} \phi(x) &= \sum_{j\in\mathbb{Z}^n} h_j \phi_{1,j}(x) \qquad (1) \\ &= \sum_{j\in\mathbb{Z}^n} h_j |\det A|^{1/2} \phi(Ax - j), \ \forall x \in \mathbb{R}^n \end{aligned}$$

where $h_j = \langle \phi, \phi_{1,j} \rangle, \forall j \in \mathbb{Z}^n$.

Given a multiresolution analysis we define, for each $i \in \mathbb{Z}$, the space W_i as the orthogonal complement of V_i in V_{i+1}: $W_i = V_{i+1} \ominus V_i$. Thus, it follows that $W_i = U_A^{-i} W_0$ and that $L^2(\mathbb{R}^n) = \bigoplus_{i\in\mathbb{Z}} W_i$.

Recalling the result of Meyer [27] that there exist $q-1$ functions $\psi^1, \ldots, \psi^{q-1}$ such that $\{\tau_j \psi^l \ : \ j \in \mathbb{Z}^n; \ l = 1, \ldots, q-1\}$ is a basis for W_0, then the set

$$\{\psi^l_{i,j} \ : \ l = 1, \ldots, q-1 \ ; i \in \mathbb{Z} \ ; j \in \mathbb{Z}^n\}$$

is a wavelet basis for $L^2(\mathbb{R}^n)$. Since $W_0 \subset V_1$, we get a dilation equation for each of the ψ^l, $l = 1, \ldots, q-1$:

$$\psi^l = \sum_{j\in\mathbb{Z}^n} g^l_j \phi_{1,j},$$

where $g^l_j = \langle \psi^l, \phi_{1,j} \rangle, \forall j \in \mathbb{Z}^n$. The coefficients h_j and g^l_j are called the *filter coefficients* of the scaling function and wavelet functions respectively.

Finding the scaling function ϕ and the functions ψ^l may be extremely difficult in general. However, in the case where the scaling functions are *characteristic functions on self-similar lattice tilings*, such basic wavelets can always be found.

2.1 Self-Similar Lattice Tilings

Let $Q, R \subset \mathbb{R}^n$ be Lebesgue measurable. Denote the characteristic function of Q by χ_Q and write $|Q|$ to denote its Lebesgue measure. Write $Q \simeq R$ if $|Q \backslash R| = |R \backslash Q| = 0$. We also state the following definitions.

Definition 24 A set Q is said to *tile* $\mathbb{R}^n$ *by integer translates* if

i) $\bigcup_{k \in \mathbb{Z}^n} (Q + k) \simeq \mathbb{R}^n$ and
ii) $Q \cap (Q + k) \simeq \emptyset$, $\forall k \in \mathbb{Z}^n \backslash \{0\}$.

It can be shown that such a set Q must have Lebesgue measure 1.

Example 25 The set $Q = [0, 1]$ tiles $\mathbb{R}$ by integer translates.

Definition 26 A set containing a unique element from each coset is called a *complete residue system.*

One can show that a complete residue system for $\mathbb{Z}^n / A\mathbb{Z}^n$ contains $q = |\det A|$ elements.

Example 27 Letting A=10 in $\mathbb{Z}$ one complete residue system for $\mathbb{Z}/A\mathbb{Z}$ is $\{0, 1, \ldots, 9\}$. This follows since each integer is equivalent to a unique element of this set, modulo 10.

For our applications, we will be particularly interested in the case of Haar wavelets. Haar wavelets are wavelets where the scaling function is the characteristic function of some set.

Definition 28 We say that a scaling function ϕ for a MRA of $L^2(\mathbb{R}^n)$ is a *Haar scaling function* if $\phi = \chi_Q$ for some measurable subset Q of $\mathbb{R}^n$.

The fact that the scaling function has to satisfy the dilation equation means that its support has to be a *self-similar tile.* In [15], Gröchenig and Madych completely characterized these scaling functions and their supports. The following theorems summarize their results.

Theorem 29 Suppose A is an acceptable dilation for $\mathbb{Z}^n$ and let Q be a measurable subset of $\mathbb{R}^n$. The function $\phi = |Q|^{-1/2} \chi_Q$ is the scaling function for a multiresolution analysis associated with A if and only if the following conditions are satisfied:

i) Q tiles $\mathbb{R}^n$ by integer translates.
ii) $AQ \simeq \bigcup_{k \in K} (Q + k)$ for some complete residue system K of $\mathbb{Z}^n / A\mathbb{Z}^n$.
iii) $Q \simeq C$ for some compact subset C of $\mathbb{R}^n$.

Since Q satisfies ii) it is called *self-similar in the affine sense*. These properties imply that the filter coefficients h_j of ϕ are identically $q^{-1/2}$ for $j \in K$ and zero otherwise. Figure 1 illustrates examples of two different self-similar tilings. The first tile (the *twin dragon* tile) is generated by the matrix and complete residue system (also called the *digit set*)

$$\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \qquad \{(0,0),(0,1)\}$$

while the second tile is generated by the matrix and digit set

$$\begin{pmatrix} 1 & 1 \\ -1 & 2 \end{pmatrix} \qquad \{(0,0),(1,0),(2,0)\}.$$

In both figures, we show nine translations of the basic tile. Since the scaling function is a characteristic function, all the non-zero coefficients in the dilation equation have the value 1. The tile is the attractor of the IFS $\{A^{-1}(x) + A^{-1}d : d \in K\}$.

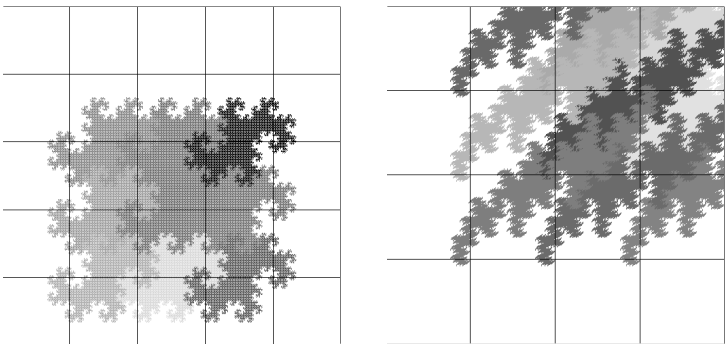


Fig. 1. Examples of self-similar tiles.

Theorem 210 Let K be a complete residue system of $\mathbb{Z}^n/A\mathbb{Z}^n$. Then there exists a unique solution of $\phi = \sum_{k\in K} q^{-1/2}\phi_{1,k}$ in $L^1(\mathbb{R}^n)$, up to multiplication by a constant. Furthermore, this solution has support in the compact set

$$Q = \left\{ \sum_{i=1}^{\infty} A^{-i}k_i \ : \ k_i \in K \right\}.$$

Theorem 211 Let A be an acceptable dilation for $\mathbb{Z}^n$ and let $Q \subset \mathbb{R}^n$. Then the function $\phi = \chi_Q$ is the scaling function of a MRA associated with A if and only if $|Q| = 1$ and Q is of the form given in Theorem 210 for some complete residue system K of $\mathbb{Z}^n/A\mathbb{Z}^n$.

Gröchenig and Madych also characterized the wavelets for such MRA.

Theorem 212 Let $\phi = \chi_Q$ be the scaling function for a MRA of $L^2(\mathbb{R}^n)$ associated with A and let $K = \{k_1, \ldots, k_q\}$ be the complete residue system generating Q. Let $U = (u_{ij})$ be a unitary $q \times q$ matrix, with $u_{1j} = q^{-1/2}$, $j = 1, \ldots, q$. For $i = 1, \ldots, q-1$ define

$$\psi^i = \sum_{j=1}^{q} u_{i+1j} \phi_{1,k_j}.$$

Then $\{\tau_j \psi^i \ : \ i = 1, \ldots, q-1 \, ; j \in \mathbb{Z}^n\}$ is a basis for W_0. Conversely, any set of basic wavelets for a MRA associated with $\phi = \chi_Q$ must arise in such a way.

An immediate corollary is that the set

$$\{\psi^l_{i,j} \ : \ l = 1, \ldots, q-1 \ ; \ i \in \mathbb{Z}; \ j \in \mathbb{Z}^n\}$$

is a basis for $L^2(\mathbb{R}^n)$.

Example 213 For $q = 2$, there are only two possible matrices U:

$$U = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ \pm 1/\sqrt{2} & \mp 1/\sqrt{2} \end{pmatrix}.$$

Another example, slightly different than the one in [15] is the following.

Example 214 For $q \geq 3$, define the unitary matrix $U = (u_{ij})$, by letting $u_{1j} = q^{-1/2}$ and

$$u_{ij} = \sqrt{\frac{2}{q}} \cos \frac{(i-1)(2j-1)\pi}{2q},$$

for $i = 2, \ldots, q$ and $j = 1, \ldots, q$.

The reader is referred to [15] for further details and examples.

2.2 Reconstruction and Decomposition Algorithm

The strength of the multiresolution analysis method lies in the reconstruction and decomposition algorithms, discovered initially by Mallat [23]. These algorithms are a fundamental component of wavelet analysis applied to signal and image processing. They will be reviewed here only briefly, mostly for notational purposes.

Suppose $(V_i)_{i \in \mathbb{Z}} \subset L^2(\mathbb{R}^n)$ is a MRA and that $f \in V_{i+1} = V_i \oplus W_i$. We have two bases: one for V_{i+1} and one for $V_i \oplus W_i$. Therefore,

$$f = \sum_{z \in \mathbb{Z}^n} s_{i+1,z}\phi_{i+1,z}$$
$$= \sum_{j \in \mathbb{Z}^n} s_{i,j}\phi_{i,j} + \sum_{l=1}^{q-1} \sum_{j \in \mathbb{Z}^n} w_{i,j}^l \psi_{i,j}^l,$$

where the *scaling coefficients* $s_{m,n} = \langle f, \phi_{m,n} \rangle$ and the *wavelet coefficients* $w_{m,n}^l = \langle f, \psi_{m,n}^l \rangle$. By using the dilation equations we obtain the identities

$$s_{i,j} = \sum_{z \in \mathbb{Z}^n} h_{z-Aj} s_{i+1,z} \quad \text{and} \quad w_{i,j}^l = \sum_{z \in \mathbb{Z}^n} g_{z-Aj}^l s_{i+1,z}, \tag{2}$$

which give the decomposition of higher resolution scaling coefficients into lower resolution scaling coefficients and wavelet coefficients. The reconstruction algorithm is given by

$$s_{i+1,z} = \sum_{j \in \mathbb{Z}^n} h_{z-Aj} s_{i,j} + \sum_{l=1}^{q-1} \sum_{j \in \mathbb{Z}^n} g_{z-Aj}^l w_{i,j}^l.$$

As is well known, this procedure yields a tree structure for the wavelet coefficients.

3 Fractal Image Compression

A fractal representation of an image tries to use self-similarity within a picture to encode the picture. The usual way of doing this in a compression scheme is to break the image up into a regular grid of large blocks (called the *source blocks* or *parent blocks*) and a regular grid of small blocks (called the *target blocks* or the *child blocks*). Then for each child block we search among all the parent blocks for the block that is the closest match. We usually allow some kind of affine modification of the image values on the block. If we represent the image as a function f then the usual transformation is of the form $\alpha f(\cdot) + \beta$.

The above "fractal block" algorithm often introduces undesirable blocking artifacts in the reconstructed image. These artifacts arise because the image is treated as a collection of disjoint child blocks. Mixed fractal-wavelet methods were introduced in an attempt to reduce these blocking artifacts.

The *Fractal-Wavelet Transform* (see [26, 29]) is based on the observation that the wavelet coefficients have a very natural tree structure, as illustrated in the following diagram (depicting the situation where the dilation is by a factor of 2):

<table>
<tr><td colspan="4">$w_{0,0}$</td></tr>
<tr><td colspan="2">$w_{1,0}$</td><td colspan="2">$w_{1,1}$</td></tr>
<tr><td>B_{20}</td><td>B_{21}</td><td>B_{22}</td><td>B_{23}</td></tr>
</table>

In this diagram, each entry B_{ij} represents a tree of infinite length. Using this notation, an example of a simple type of Fractal-Wavelet Transform is the transformation W defined by:

$$W_1 : B_{10} \to B_{20}, \quad W_2 : B_{11} \to B_{21},$$

and

$$W_3 : B_{10} \to B_{22}, \quad W_4 : B_{11} \to B_{23},$$

with associated multipliers α_i, $1 \leq i \leq 4$. Diagrammatically, this IFSW transforms B_{00} into

$w_{0,0}$			
$w_{1,0}$		$w_{1,1}$	
$\alpha_1 B_{10}$	$\alpha_2 B_{11}$	$\alpha_3 B_{10}$	$\alpha_4 B_{11}$

where $|\alpha_i| < 1/\sqrt{2}$. The restrictions on the α_i follow from the condition that the wavelet coefficient sequence $w_{i,j}$ is square summable.

The map W has a unique fixed point function $\bar{u}$ whose wavelet coefficients decay geometrically according to the α_i's.

All the previous work on fractal-wavelet methods have used separable wavelets. This was mainly due to the lack of an algorithm to do the wavelet transform for non-separable wavelets. In this paper, we concentrate on using non-separable wavelets.

4 Periodic Non-Separable Wavelet Transform

In this section, we present a brief discussion of the algorithm for the discrete periodic wavelet transform for non-separable wavelets (for a more complete description and discussion of the algorithm see the paper [24]). Our algorithm is general in the sense that you can use *any* valid scaling function filter and wavelet filters for an appropriate dilation matrix. However, in our applications we concentrate on the case of Haar wavelets. We do this mainly because very few wavelet filters for smooth non-separable wavelets are known ([2]).

For simplicity, we restrict ourselves to two dimensions. The extension to higher dimensions is reasonably straightforward.

We assume that we are given a dilation matrix A and that our signal is of size $det(A)^N$ by $det(A)^N$ for some N. Furthermore, we make the assumption that

$$trace(A) = 0 \bmod det(A) \tag{3}$$

and that the only solution to $A(a,0)^T = (0,0)^T$ is $a = 0$. It is possible to weaken these assumptions, but the algorithm is slightly more complicated.

The basic idea is that (2) tells us how to filter and downsample our given signal, so we implement these equations. We use the proper arithmetic on the subscripts to keep track of where to subsample and how to convolve. However, this is not entirely straightforward.

First, the data is downsampled by a factor of $det(A)$ on each iteration of the wavelet analysis algorithm. This may make the storage and indexing of the data complicated. For example, if the image is square and $det(A) = 2$, then after one iteration you no longer have a square. This also complicates the periodization (performing the arithmetic correctly). The indexing problem can be solved by using A to help with the indexing from one "level" to the next. The periodization problem (performing the correct arithmetic) can be solved by noting that after one subsampling we are on the sublattice $A(\Lambda) \cong \Lambda/ker(A)$, where $\Lambda = \mathbb{Z}_{det(A)^N} \times \mathbb{Z}_{det(A)^N}$ is the lattice (module) of integers modulo $det(A)^N$. Thus, we perform our arithmetic modulo $det(A)^N$ and then reduce modulo $ker(A)$.

This leads to the second problem. In general, the subsampling lattices $A^n(\Lambda)$ get more complicated with each iteration. Thus, in order to perform the correct arithmetic on the n^{th} stage it would be necessary to reduce modulo $ker(A^n)$. It would be necessary to know $ker(A^n)$ for all n to do this and this reduction would make the algorithm very slow.

Finally, suppose A has an eigenvector v over $\mathbb{Z}_{det(A)}$ associated with an eigenvalue λ which is invertible modulo $det(A)$. Then $det(A)^{N-1}v$ is an eigenvector of A over Λ with eigenvalue λ. So after repeatedly sub-sampling the lattice Λ using A, we will eventually get to a stage where our sub-sampling points are no longer changing; the wavelet analysis algorithm cannot be completed. The only way around this problem is to restrict our possible dilation matrices to those that have no such eigenvectors. It is easy to see that if A has an invertible eigenvalue modulo $det(A)^N$ then A must necessarily have an invertible eigenvalue modulo $det(A)$. Thus it is sufficient to check for eigenvalues modulo $det(A)$.

Our assumption given in equation (3) eliminates these problems. Since A is a 2×2 matrix, the characteristic polynomial of A is

$$\lambda^2 - trace(A)\lambda + det(A).$$

If $trace(A) = 0 \bmod det(A)$, then A has no non-invertible eigenvalues modulo $det(A)$, and thus has no associated eigenvectors. Furthermore, by the Cayley-Hamilton Theorem (which remains true for matrices over commutative rings, see [4]) we have

$$A^2 \equiv 0 \bmod det(A)$$

which means that each entry of A^2 is a multiple of $det(A)$ so that

$$A^2(\Lambda) \subset \det(A)\Lambda.$$

In fact, this equation is an equality. Thus, while $A(\Lambda)$ might be some arbitrary submodule of Λ, $A^2(\Lambda)$ sits in Λ very simply. In order to do arithmetic in $A^2(\Lambda)$ we simply perform our arithmetic modulo $det(A)^{N-1}$ in each coordinate! Thus, we can go from our "square" arrays of data to "rectangular" and then back to "square".

Here is an example to illustrate these ideas.

Example 41 Let A be the dilation matrix

$$A = \begin{pmatrix} 2 & 1 \\ 1 & 4 \end{pmatrix}.$$

Then $A(2,1)^T \equiv 6(2,1)^T \bmod 7$ so A has an eigenvector modulo 7 corresponding to the invertible eigenvalue 6.

In Figure 1 we use the lattice $\Lambda = \mathbb{Z}_{49} \times \mathbb{Z}_{49}$ and show $A(\Lambda)$ and $A^2(\Lambda)$. For this example $A^n(\Lambda) = A^2(\Lambda)$ for $n \geq 2$.

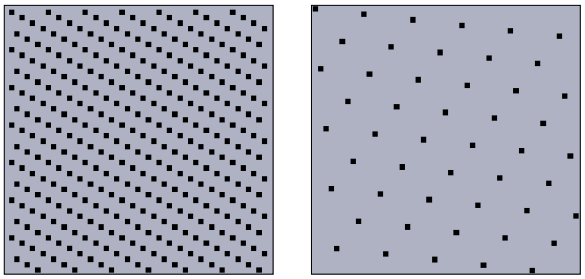


Fig. 2. $A(\Lambda)$ and $A^n(\Lambda)$ for $n \geq 2$

Application to Image Compression

The motivation for using non-separable wavelets in image processing is that any "blocking" artifacts created by these wavelets would be less noticeable to the human eye. This is especially true if we use Haar wavelets. In this case, using separable wavelets creates very noticeable blockiness in the image since separable Haar wavelets are characteristic functions of squares. When we use non-separable wavelets, this blockiness is considerably reduced.

Figure 3 illustrates the basic difference between the use of separable and non-separable Haar wavelets. The first image uses the separable Haar wavelets and represents truncating the bottom 2 levels of the wavelet tree while the second uses Haar wavelets based on the twin dragon tiling (the first tiling in figure 1) and represents truncating the bottom 4 levels. The information content is similar in the two pictures since for the separable Haar wavelets each level represents downsampling by a factor of 4 ($det(A) = 4$ since $A = 2I$) while for the non-separable one each level represents downsampling by a factor of 2.

In figure 4 we illustrate the use of Haar wavelets based on the second tiling in figure 1. In this case we have truncated the finest three levels of the wavelet tree (which corresponds to subsampling by a factor of nine, since $det(A) = 3$). The second image in figure 4 is the result of using the wavelet analysis and synthesis

algorithm for wavelets whose dilation matrix A does not satisfy the assumption in equation (3). This matrix is the same one whose subsampling lattice is shown in figure 2.

Figure 5 shows the result of two types of encoding: zero-tree and fractal-wavelet. First we performed a wavelet decomposition using the twin dragon Haar wavelets. Next we traversed the wavelet tree truncating the tree when the total energy of the branch was less than some specified threshold. The image on the left is the image reconstructed from this pruned wavelet tree. The image on the right shows the result of fractal-wavelet compression using the twin dragon Haar wavelets. We first computed the wavelet tree for the lenna image and store the first 12 levels of the tree. Next we used these levels to predict the higher resolution levels in the manner described in Section 3.

Fig. 3. Left: Separable Haar 4:1. Right: Twin-dragon Haar 4:1.

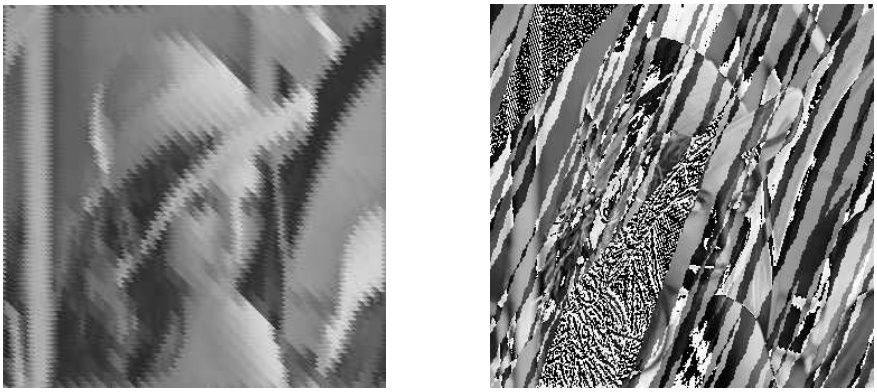


Fig. 4. Left: Non-separable Haar. Right: Improper subsampling.

Fig. 5. Left: Zero-tree encoding. Right: Fractal-Wavelet encoding.

5 Complex Bases

Now we turn to a discussion of the wavelet transform algorithm for Haar wavelets based on the connection between self-similar lattice tilings and complex bases. In order to do this, we first give some background on complex bases. We use a base and digits from an integer set to represent the numbers in a given system. For more on the theory of complex bases, we refer the reader to some papers by Gilbert [10–12]. Most of the details for the results that follow can be found there.

5.1 Background

The concept of representing numbers in a base is a very simple and familiar one. In real numbers, the most common examples are the decimal and binary systems.

Example 51 Representing positive real numbers with $\mathbb{N}$ as the integers.
Base 10, digits $\{0, 1, 2, \ldots, 9\}$

$$193_{10} = 1 \cdot 10^2 + 9 \cdot 10^1 + 3 \cdot 10^0.$$

Base 2, digits $\{0, 1\}$

$$11010_2 = 1 \cdot 2^4 + 1 \cdot 2^3 + 0 \cdot 2^2 + 1 \cdot 2^1 + 0 \cdot 2^0 = 26_{10}.$$

The same idea carries over to complex numbers. In this case, the integers are the Gaussian integers:

Definition 52 The set of *Gaussian integers*, denoted by $\mathbb{Z}[i]$, is the set of complex numbers of the form $a+bi$, where $a, b \in \mathbb{Z}$. The *norm* of a Gaussian integer $a+bi$ is defined as

$$\text{norm}(a+bi) \equiv a^2 + b^2.$$

Definition 53 A *valid base* is a pair (b, D) where b is a Gaussian integer and $D \subset \mathbb{Z}[i]$, such that $0 \in D$ and every integer z can be represented uniquely as a sum of powers of b, with coefficients in D. More precisely, each $z \in \mathbb{Z}[i]$ can be written uniquely as

$$z = \sum_{j=0}^{t} a_j b^j, \text{ where } a_j \in D \text{ and } t \in \mathbb{N}.$$

If z has this form, write $z = (a_t a_{t-1} \cdots a_1 a_0)_b$. This is called *base b positional notation*. The integer b is often referred to as the *base*, and the set D is called the *digit set*. When the base is understood, the subscript b in the positional notation is often omitted.

A result below will show that $(-1+i, \{0,1\})$ is a valid base.

Example 54 We can expand 3 in $(-1+i, \{0,1\})$:

$$3 = (-1+i)^3 + (-1+i)^2 + (-1+i)^0 = (1101)_{-1+i}.$$

Definition 55 Let (b, D) be a valid base. If $z \in \mathbb{C}$ has the form

$$z = \sum_{j=-\infty}^{t} a_j b^j, \text{ where } a_j \in D \text{ and } t \in \mathbb{N},$$

then the *radix expansion of z in base (b, D)* is defined as

$$z = (a_t a_{t-1} \cdots a_1 a_0 \,.\, a_{-1} a_{-2} \cdots)_b.$$

The point between the digits a_0 and a_{-1} is called the *radix point*. The string of digits to the left of the radix point is called the *integer part* of z, while the string to the right is called the *radix part*. Another name for the radix expansion of a complex number is the *address* of the number.

Example 56 The complex number $(-1-8i)/5$ can be expressed in base $(-1+i, \{0,1\})$ as

$$(-1-8i)/5 = (111.\overline{10})_{-1+i}.$$

where $\overline{10}$ indicates that the string 10 is repeated indefinitely. This can be seen by considering a geometric series in $(-1+i)^{-1}$.

Proposition 57 (Gilbert [11]) If (b, D) is a valid base, then D is a complete residue system for $\mathbb{Z}[i]$ modulo b and hence contains $\text{norm}(b)$ elements.

Theorem 58 (Gilbert [13]) Each $z \in \mathbb{C}$ has an infinite radix expansion in a valid base. However, this expansion may not necessarily be unique.

It therefore makes sense to define the *fundamental tile* of a valid base.

Definition 59 Given a valid base (b, D), define the *fundamental tile* $T(b, D)$ as the set of complex numbers with zero integer part in the base.

By Theorem 58, $\mathbb{C} = \cup_{z \in \mathbb{Z}[i]}(T(b, D) + z)$. The following result shows that there are many valid bases.

Theorem 510 (Davio, Deschamps and Gossart [7]) Given any $b \in \mathbb{Z}[i]$ with modulus larger than one, except 2 and $1 \pm i$, there exists a complete residue system D such that (b, D) is a valid base for $\mathbb{C}$.

The following result gives an entire class of complex bases.

Theorem 511 (Kátai and Szabó [18]) If $b \in \mathbb{Z}[i]$, with norm N and $D = \{0, 1, 2, \ldots, N-1\}$, then (b, D) is a valid base for $\mathbb{C}$ iff $b = -n \pm i$, for some positive integer n.

Further generalizations can be found in [9, 17].

Corollary 512 The pair $(-1 + i, \{0, 1\})$ is a valid base.

5.2 Representation in a Complex Base

There are various algorithms for determining the representation of Gaussian integers in a valid base. These are due to Gilbert [9, 12–14].

Algorithm 513 (Division Algorithm) (Gilbert [13]) Let (b, D) be a valid base. Since D is a complete residue system for $\mathbb{Z}[i]$ modulo b then, given $z \in \mathbb{Z}[i]$, there exist unique integers $q_j \in \mathbb{Z}[i]$ and $a_j \in D$, $j = 1 \ldots t$, $t \in \mathbb{N}^+$ such that

$$\begin{aligned} z &= q_1 b + a_0 \\ q_1 &= q_2 b + a_1 \\ &\vdots \\ q_t &= 0b + a_t. \end{aligned}$$

Hence, $z = (a_t \ldots a_1 a_0)_b$.

Example 514 Let $z = 5 + 12i$. Find the address of z in $(-2 + i, \{0, 1, 2, 3, 4\})$.

Using the Division Algorithm

$$\begin{aligned} 5 + 12i &= (2 - 5i)(-2 + i) + 4 \\ 2 - 5i &= (-1 + 2i)(-2 + i) + 2 \\ -1 + 2i &= 2(-2 + i) + 3 \\ 2 &= 0(-2 + i) + 2 \end{aligned}$$

Hence the address of $5 + 12i$ is (2324).

A fast algorithm, called the Clearing Algorithm, also exists for finding the expansion of integers in bases of the form (b, D) where $D \subset \mathbb{Z}$. For simplicity, we consider only bases $(-n+i, \{0, 1, \ldots, n^2\})$, $n \in \mathbb{N}^+$. The reader is referred to [9] for the general result.

Algorithm 515 (Clearing Algorithm) (Gilbert [9]) Consider the valid base $(-n + i, \{0, 1, \ldots, n^2\})$ and let $p(x)$ be the minimal polynomial of $b = -n + i$. Thus

$$p(x) = x^2 + 2nx + n^2 + 1.$$

Then the representation of any integer $z \in \mathbb{Z}[i]$ in the base can be obtained as follows.

Begin with $z = m(b) = a_k b^k + \cdots + a_1 b + a_0$, an expansion of z in powers of b with integer coefficients. For instance, any Gaussian integer $c + id$ can be expanded in powers of $b = -n + i$ as $c + id = db + (c + nd)$. Consider this expansion as an element $m(b)$ of the polynomial ring $\mathbb{Z}[b]$. Let r be the least integer such that $a_r \notin D$. If no such r exists, then $m(b)$ is the unique expansion of z in the base. We call such a polynomial *clear*.

If r exists, then let s be the integer such that $0 \leq a_r + s(n^2+1) \leq n^2$. Add sb^r times $p(b)$ to $m(b)$. Remember that we perform this operation in the polynomial ring $\mathbb{Z}[b]$. Call this new polynomial $m_1(b)$.

However, $p(-n + i) = 0$, thus $m(b)$ and $m_1(b)$ are equal in $\mathbb{C}$. Hence, $m_1(b)$ is an expansion of z in $\mathbb{Z}[b]$. In addition, the coefficient of the r-th power of b in $m_1(b)$ is a digit. We say that the a_r has been *cleared*.

Repeat the process of clearing coefficients, by induction on r, until a clear polynomial is obtained. The resulting polynomial is the expansion of z in (b, D). This process must terminate after a finite number of steps.

Example 516 Determine the expansion of $5+12i$ in base $(-2+i, \{0, 1, 2, 3, 4\})$.

The minimal polynomial of $b = -2 + i$ is $x^2 + 4x + 5$. Hence, $b^2 + 4b + 5 = 0$ and, by abuse of notation, we can write this as $(1\ 4\ 5)_b = 0$. (Note that 5 is not in the digit set.)

Begin with the expansion $5 + 12i = 12b + 29 = (12\ 29)_b$. Then, we clear the polynomial in $\mathbb{Z}[b]$ as follows:

$$\begin{array}{rrrr} & & 12 & 29 \\ & -5 & -20 & -25 \\ \hline & -5 & -8 & 4 \\ 2 & 8 & 10 & \\ \hline 2 & 3 & 2 & 4 \end{array}$$

Hence, the expansion of $5+12i$ in base $(-2+i, \{0,1,2,3,4\})$ is (2324). A quick calculation verifies that this is indeed correct.

6 Linking Complex Bases to Wavelets

6.1 Addressing

The theory of MRA and complex bases are related. Multiplication by the base $b = c + id$ in $\mathbb{C}$ is equivalent to multiplication by the acceptable dilation

$$A = \begin{pmatrix} c & -d \\ d & c \end{pmatrix}$$

in $\mathbb{R}^2$, via the natural association $s + it \leftrightarrow (s,t)^T$. Here, $\det A = c^2 + d^2 = \text{norm}(b)$ and $\lambda = c \pm id$. The fundamental tile $T(b,D)$ for a valid base (b,D) is simply the basic tile Q of A, where the complete residue system K is identified with D through the natural association. Therefore,

$$T(b,D) = \left\{ \sum_{j=-\infty}^{-1} a_j b^j \ : \ a_j \in D \right\} \equiv \left\{ \sum_{i=1}^{\infty} A^{-i} k_i \ : \ k_i \in K \right\} = Q.$$

6.2 Reconstruction and Decomposition Algorithm

Consider the reconstruction and decomposition algorithms for a Haar MRA associated with an acceptable dilation A for $\mathbb{Z}^2$, with complete residue system K. Assume A is associated with a complex base (b,D) in the sense of Section 6.1. Rewriting the equation for the wavelet decomposition, we have

$$\begin{aligned} s_{i,j} &= \sum_{z \in \mathbb{Z}^2} h_{z-Aj} s_{i+1,z} \\ &= \sum_{k \in K} h_k s_{i+1,Aj+k}, \end{aligned}$$

since $h_k = 0$ when $k \notin K$. We can translate this identity to the language of complex bases as

$$s_{i,j} = \sum_{d \in D} h_d s_{i+1,bj+d},$$

where the scaling filters and coefficients are re-indexed, via the natural association given above.

By looking at the decomposition algorithm in the language of complex bases, we can consider $bj + d$ as the address of j shifted to the left by one place, with d added as the zero-th order digit. We can then think of the index i as the length of the address of j. The scaling coefficient associated with the point j is thus a linear combination of the coefficients of the points whose addresses are one digit longer than that of j, and start with the address of j. In other words, letting j have address $(a_t \dots a_1)$ in base b we have

$$s_{t,(a_t \dots a_1)} = \sum_{d \in D} h_d s_{t+1,(a_t \dots a_1 d)}.$$

This gives a precise and **implementable** tree structure for the coefficients. A similar relation holds for the reconstruction algorithm.

6.3 Application to Images

We present here an example of this process applied to images.

Example 61 Perform the decomposition of the function f in base $(-1+i, \{0,1\})$, where f is the *simple function* given by the following diagram:

1	2
0	1

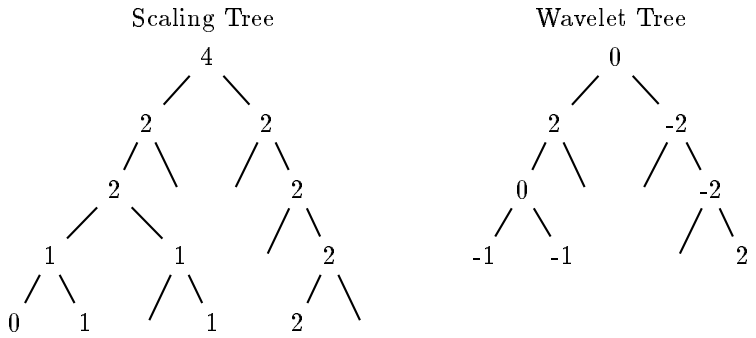
By this representation, we take the boxes to be unit squares in $\mathbb{C}$, with the left hand corner at 0. The value of the function on each square is the number written inside the square.

The addresses of the values are

Integer	Address	Value
0	0000	0
1	0001	1
i	0011	1
1+i	1110	2

The decomposition begins at level 4, the length of the longest address.

Choose the standard Haar wavelet basis as given in Example 213. We then obtain the scaling and wavelet trees using the decomposition algorithm. The lowest level on the scaling tree represents the initial values of the function. Going down the left branch in a tree represents the digit 0. The right represents the digit 1. For example, the value of the function on the square i, which is at address 0011, is placed in the node down the scaling tree left, left, right, right. We normalize the scaling filters in this example for the purposes of clarity. The correct tree would reduce the values by $\sqrt{2}$ at each level, beginning with level 3. Empty nodes are set to zero.



Notice the number of empty nodes in the trees. This demonstrate the asymmetrical nature of the trees generated by this process. This occurs since the support of the original function sits inside the principal tile, but is not the entire tile; it is a consequence of the zero padding of the function.

Example 62 An illustration of the above method applied to a 512×512 image of Lena is shown in Figure 6. The wavelet decomposition was performed on this image, using the complex base $(2+i, \{0, 1, i, -i, -2-3i\})$. The wavelets are given by the 5×5 unitary matrix of Example 214. A threshold of 50 was used to set individual wavelet coefficients to zero. In this case, 94% of the coefficients were set to zero. The tile of this base is disconnected and dust-like. It is shown in Figure 7 along with the first basic wavelet.

Fig. 6. Wavelet compression on dust. Left: Original 512×512 of Lena. Right: Reconstruction of Lena from pruned wavelet tree. PSNR=27.22 dB.

The implementation of compression methods on wavelet trees of complex bases might not be trivial. Current compression methods, such as the zero-tree

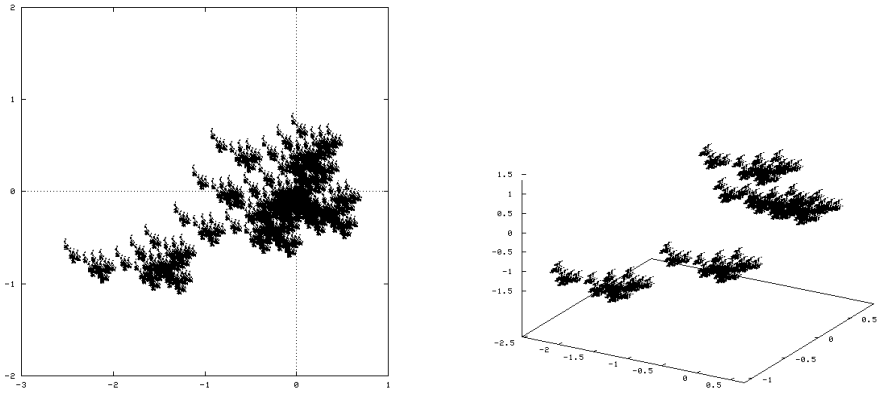


Fig. 7. Fundamental tile of the base $(2+i, \{0, 1, i, -i, -2-3i\})$ with a basic wavelet.

(see [3]), rely on extremely symmetric and balanced distributions of the coefficients in the expansion tree. The trees created by complex bases are tremendously unbalanced, as is seen in Example 61, above. This asymmetry is much greater with larger numbers of points such as in images. There may also be fractal-wavelet transforms (see [26, 29]) which could utilize the structure of the complex bases. This could lead to a general fractal-wavelet transform for non-separable wavelets.

It is possible that no method can be constructed which will allow compression to be viable for these wavelets. If this is the case one could investigate compression on translations of the image which result in more symmetric trees. It may even be necessary to assume that the image is exactly the tile by mapping it onto the tile of an appropriate size. Computational complexity of such methods may result in algorithms which are slower than the standard ones.

7 Conclusion

The two methods given here for non-separable wavelet transforms are seen to work well in their respective domains. While these domains have significant overlap, neither algorithm supersedes the other; they both have strengths and weaknesses.

Acknowledgment. The first author was supported in part by a Natural Sciences and Engineering Research Council of Canada (NSERC) Collaborative Grant (under C. Tricot and E.R. Vrscay) in the form of a Postdoctoral Fellowship. The support of an NSERC Postgraduate Scholarship is gratefully acknowledged by the second author. The authors also thank Professors William Gilbert and Edward Vrscay, as well as the members of the Waterloo "fractalators" (`http://links.uwaterloo.ca`) for their most helpful advice and discussions.

References

1. M. F. Barnsley, *Fractals Everywhere*, Academic Press, New York, 1988.
2. E. Belogay and Yang Wang, Arbitrarily smooth orthogonal nonseparable wavelets in $\mathbb{R}^2$, to appear in *SIAM Math. Analysis.*
3. J. J. Benedetto and M. W. Frazier, editors. *Wavelets: Mathematics and Applications.* Studies in Advanced Mathematics. CRC Press, 1994.
4. William C. Brown, *Matrices over Commutative Rings*, Marcel Dekker, New York, 1993.
5. C. Cabrelli, C. Heil, and U. Molter, Accuracy of lattice translates of several multidimensional refinable functions, to appear in *J. Approximation Theory.*
6. I. Daubechies, *Ten Lectures on Wavelets*, SIAM Press, Philadelphia, PA, 1992.
7. M. Davio, J. P. Deschamps and C. Gossart. *Complex Arithmetic.* Philips MBLE Research Lab. Report R369, May 1978.
8. Yuval Fisher , *Fractal Image Compression: Theory and Appli cations* Springer-Verlag, New York, 1995.
9. W. Gilbert. Radix representations of quadratic fields. *J. Math. Anal. Appl.*, 83:264–274, 1981.
10. W. Gilbert. Fractal geometry derived from complex bases. *Math. Intelligencer*, 4(2):78–86, 1982.
11. W. Gilbert. Geometry of radix representations. In C. Davis, B. Grünbaum and F. A. Sherk, editors, *The Geometric Vein, The Coxeter Festschrift*, pages 129–139. Springer-Verlag, 1982.
12. W. Gilbert. Arithmetic in complex bases. *Mathematics Magazine*, 57(2):77–81, 1984.
13. W. Gilbert. Complex based number systems. Unpublished, May 1994.
14. W. Gilbert. The division algorithm in complex bases. *Canadian Mathematical Bulletin*, 39(1):47–54, 1996.
15. K. Gröchenig and W. R. Madych. Multiresolution analysis, haar bases, and self-similar tilings of R^n. *IEEE Trans. Inform. Theory*, 39:556–568, 1992.
16. J. E. Hutchinson, Fractals and Self-similarity, *Indiana Univ. Math. J.* **30** (1981), 713-747.
17. I. Kátai and B. Kovács. Canonical number systems in imaginary quadratic fields. *Acta Math. Acad. Sci. Hungaricae*, 37:159–164, 1981.
18. I. Kátai and J. Szabó. Canonical number systems for complex integers. *Acta. Sci. Math. (Szeged)*, 37:255–260, 1975.
19. J. Kovačević and M. Vetterli, Nonseparable two- and three-dimensional wavelets, *IEEE Trans. on Signal Processing* **43** (1995), 1269–1273.
20. J. C. Lagarias and Yang Wang, Haar-type orthonormal wavelet basis in $\mathbb{R}^2$, *J. Fourier Analysis and Appl.*, **2** (1995), 1-14.
21. Kevin Leeds, Dilation Equations with Matrix Dilations, PhD Thesis, Department of Math, Georgia Tech, 1997.
22. N. Liu, *Fractal Imaging*, Academic Press, 1997.
23. S. Mallat. Multiresolution approximation and wavelet orthonormal bases of $L^2(\mathbb{R})$. *Trans. AMS*, 315:69–88, 1989.
24. F. Mendivil, A discrete periodic wavelet transform for non-separable wavelets, preprint, 1998.
25. F. Mendivil. The application of a fast non-separable discrete periodic wavelet transform to fractal image compression, preprint, 1998. *Submitted to Fractals in Engineering 99.*

26. F. Mendivil and E. R. Vrscay. Correspondence between fractal-wavelet transforms and iterated function systems with grey level maps. In E. Lutton, C. Tricot and J. Lévy Véhel, editors, *L'Ingénieur et les fractales*, pages 54–64. Springer Verlag, 1997.
27. Y. Meyer. *Ondelettes, fonctions splines et analyses graduées.* Rapport CEREMADE n.8703, 1987.
28. C. A. Micchelli and H. Prautzsch, Uniform refinement of curves, Linear Algebra Appl., **114/115** (1989), pp. 841-870.
29. E. R. Vrscay. A new class of fractal-wavelet transforms for image representation and compression. *Can. J. Elect. Comp. Eng.*, 23:69–83, 1998.
30. G. Strang and T. Nguyen, *Wavelets and Filter Banks*, Wellesley-Cambridge Press, Wellesley, 1996.
31. Yang Wang, Self-Affine Tiles, preprint, to appear in *Proc. Chinese Univ. of Hong Kong Workshop on Wavelet, 1997.*