The Finite Ridgelet Transform for Image Representation

Minh N. Do, *Member, IEEE,* and Martin Vetterli, *Fellow, IEEE*

Abstract— **The ridgelet transform (Candès and Donoho, 1999) was introduced as a sparse expansion for functions on continuous spaces that are smooth away from discontinuities along lines. In this paper, we propose an orthonormal version of the ridgelet transform for discrete and finite-size images. Our construction uses the finite Radon transform (FRAT) (Bolker, 1987; Matúš and Flusser, 1993) as a building block. To overcome the periodization effect of a finite transform, we introduce a novel ordering of the FRAT coefficients. Taking the one-dimensional wavelet transform on the projections of the FRAT in a special way results in the finite ridgelet transform (FRIT), which is invertible, non-redundant, and computed via fast algorithms. Furthermore, our construction leads to a family of directional and orthonormal bases for images. Numerical results show that the FRIT is more effective than the wavelet transform in approximating and denoising images with straight edges.**

Keywords— **wavelets, ridgelets, Radon transform, directional bases, discrete transforms, non-linear approximation, image representation, image denoising.**

I. Introduction

Many image processing tasks take advantage of *sparse* representations of image data where most information is packed into a small number of samples. Typically, these representations are achieved via invertible and non-redundant transforms. Currently, the most popular choices for this purpose are the wavelet transform [1], [2], [3] and the discrete cosine transform [4].

The success of wavelets is mainly due to the good performance for piecewise smooth functions in one dimension. Unfortunately, such is not the case in two dimensions. In essence, wavelets are good at catching zero-dimensional or *point* singularities, but two-dimensional piecewise smooth signals resembling images have one-dimensional singularities. That is, smooth regions are separated by edges, and while edges are discontinuous across, they are typically smooth curves. Intuitively, wavelets in two dimensions are obtained by a tensor-product of one dimensional wavelets and they are thus good at isolating the discontinuity across an edge, but will not see the smoothness along the edge.

To overcome the weakness of wavelets in higher dimensions, Candès and Donoho [5], [6] recently pioneered a new system of representations named *ridgelets* which deal effectively with *line* singularities in 2-D. The idea is to map a line singularity into a point singularity using the Radon transform [7]. Then, the wavelet transform can be used to effectively handle the point singularity in the Radon domain. Their initial proposal was intended for functions defined in the *continuous* $\mathbb{R}^2$ space.

For practical applications, the development of *discrete* versions of the ridgelet transform that lead to algorithmic implementations is a challenging problem. Due to the radial nature of ridgelets, straightforward implementations based on discretization of continuous formulae would require interpolation in polar coordinates, and thus result in transforms that would be either redundant or can not be perfectly reconstructed.

In [8], [9], [10], the authors take the redundant approach in defining discrete Radon transforms that can lead to invertible discrete ridgelet transforms with some appealing properties. For example, a recent preprint [10] proposes a new notion of Radon transform for data in a rectangular coordinate such that the lines exhibit geometrical faithfulness. Their transform is invertible with a factor four oversampled. However, the inverse transform is ill-conditioned in the presence of noise and requires an iterative approximation algorithm.

In this paper, we propose a discrete ridgelet transform that achieves both invertibility and non-redundancy. In fact, our construction leads to a large family of *orthonormal* and directional bases for digital images, including adaptive schemes. As a result, the inverse transform is numerically stable and uses the same algorithm as the forward transform. Because a basic building block in our construction is the finite Radon transform [11], which has a wrap-around (or aliased line) effect, our ridgelet transform is not geometrically faithful. The properties of the new transform are demonstrated and studied in several applications.

As an illustration, consider the image denoising problem where there exist other approaches that explore the geometrical regularity of edges, for example by chaining adjacent wavelet coefficients and then thresholding them over those contours [12]. However, the discrete ridgelet transform approach, with its "built-in" linear geometrical structure, provide a more direct way – by simply thresholding significant ridgelet coefficients – in denoising images with straight edges.

The outline of this paper is as follows. In the next section we review the concept and motivation of ridgelets in the continuous domain. In Section III, we introduce the finite Radon transform with a novel ordering of coefficients as a key step in our discrete ridgelet construction. The

M. N. Do was with the Audiovisual Communications Laboratory, Department of Communication Systems, Swiss Federal Institute of Technology, Lausanne, Switzerland. He is now with the Department of Electrical and Computer Engineering and the Beckman Institute, University of Illinois, Urbana IL 61801 (email: minhdo@uiuc.edu).

M. Vetterli is with the Audiovisual Communications Laboratory, Department of Communication Systems, Swiss Federal Institute of Technology, Lausanne, Switzerland, and with the Department of Electrical Engineering and Computer Science, University of California, Berkeley CA 94720 (email: martin.vetterli@epfl.ch).

This work was supported in part by a Ph.D. Fellowship from the Department of Communication Systems, EPFL, and the Swiss National Science Foundation under Grant 21-52439.97.

finite Radon transform is then studied within the theory of frames. The finite ridgelet transform is defined in Section IV, where the main result is a general family of orthonormal transforms for digital images. In Sections V, we propose several variations on the initial design of the finite ridgelet transform. Numerical experiments are presented in Section VI, where the new transform is compared with the traditional ones, especially the wavelet transform. We conclude in Section VII with some discussions and an outlook.

II. Continuous Ridgelet Transform

We start by briefly reviewing the ridgelet transform and showing its connections with other transforms in the continuous domain. Given an integrable bivariate function $f(\boldsymbol{x})$, its *continuous ridgelet transform* (CRT) in $\mathbb{R}^2$ is defined by [5], [6]

$$CRT_f(a,b,\theta) = \int_{\mathbb{R}^2} \psi_{a,b,\theta}(\boldsymbol{x}) f(\boldsymbol{x}) d\boldsymbol{x}, \qquad (1)$$

where the ridgelets $\psi_{a,b,\theta}(\boldsymbol{x})$ in 2-D are defined from a wavelet-type function in 1-D $\psi(x)$ as

$$\psi_{a,b,\theta}(\boldsymbol{x}) = a^{-1/2}\psi((x_1\cos\theta + x_2\sin\theta - b)/a). \qquad (2)$$

Figure 1 shows an example ridgelet function, which is oriented at an angle θ and is constant along the lines $x_1\cos\theta + x_2\sin\theta = const.$

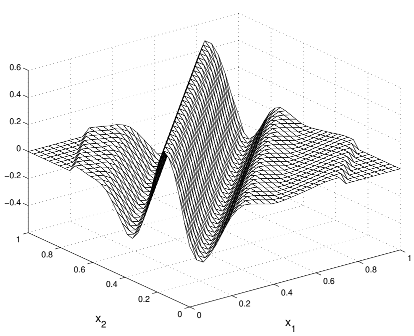


Fig. 1. An example ridgelet function $\psi_{a,b,\theta}(x_1, x_2)$.

For comparison, the (separable) continuous wavelet transform (CWT) in $\mathbb{R}^2$ of $f(\boldsymbol{x})$ can be written as

$$CWT_f(a_1,a_2,b_1,b_2) = \int_{\mathbb{R}^2} \psi_{a_1,a_2,b_1,b_2}(\boldsymbol{x}) f(\boldsymbol{x}) d\boldsymbol{x}, \qquad (3)$$

where the wavelets in 2-D are tensor products

$$\psi_{a_1,a_2,b_1,b_2}(\boldsymbol{x}) = \psi_{a_1,b_1}(x_1)\psi_{a_2,b_2}(x_2), \qquad (4)$$

of 1-D wavelets, $\psi_{a,b}(t) = a^{-1/2}\psi((t-b)/a)$.[1]

As can be seen, the CRT is similar to the 2-D CWT except that the *point* parameters (b_1, b_2) are replaced by the *line* parameters (b, θ). In other words, these 2-D multiscale transforms are related by:

Wavelets:	$\rightarrow$	$\psi_{scale,\ point-position}$,
Ridgelets:	$\rightarrow$	$\psi_{scale,\ line-position}$.

As a consequence, wavelets are very effective in representing objects with isolated point singularities, while ridgelets are very effective in representing objects with singularities along lines. In fact, one can think of ridgelets as a way of concatenating 1-D wavelets along lines. Hence the motivation for using ridgelets in image processing tasks is appealing since singularities are often joined together along edges or contours in images.

In 2-D, points and lines are related via the Radon transform, thus the wavelet and ridgelet transforms are linked via the Radon transform. More precisely, denote the Radon transform as

$$R_f(\theta,t) = \int_{\mathbb{R}^2} f(\boldsymbol{x})\delta(x_1\cos\theta + x_2\sin\theta - t) d\boldsymbol{x}, \qquad (5)$$

then the ridgelet transform is the application of a 1-D wavelet transform to the slices (also referred to as projections) of the Radon transform,

$$CRT_f(a,b,\theta) = \int_{\mathbb{R}} \psi_{a,b}(t) R_f(\theta,t) dt. \qquad (6)$$

It is instructive to note that if in (6) instead of taking a 1-D wavelet transform, the application of a 1-D Fourier transform along t would result in the 2-D Fourier transform. More specifically, let $F_f(\boldsymbol{\omega})$ be the 2-D Fourier transform of $f(\boldsymbol{x})$, then we have

$$F_f(\xi\cos\theta, \xi\sin\theta) = \int_{\mathbb{R}} e^{-j\xi t} R_f(\theta,t) dt. \qquad (7)$$

This is the famous *projection-slice* theorem and is commonly used in image reconstruction from projection methods [13], [14]. The relations between the various transforms are depicted in Figure 2.

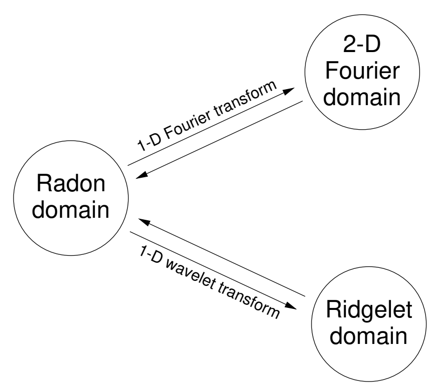


Fig. 2. Relations between transforms. The ridgelet transform is the application of 1-D wavelet transform to the slices of the Radon transform, while the 2-D Fourier transform is the application of 1-D Fourier transform to those Radon slices.

[1] In practice, however one typically enforces the same dilation scale on both directions thus leading to three wavelets corresponding to horizontal, vertical and diagonal directions.

III. Finite Radon Transform

A. Forward and Inverse Transforms

As suggested in the previous section, a discrete ridgelet transform can be constructed using a discrete Radon transform. Numerous discretizations of the Radon transforms have been devised to approximate the continuous formulae [15], [13], [14], [16], [17], [18]. However, most of them were not designed to be invertible transforms for digital images. Alternatively, the *finite* Radon transform theory (which means transform for *finite* length signals) [11], [19], [20], [21] originated from combinatorics, provides an interesting solution. Also, in [22], a closely related transform is derived from the periodization of the continuous Radon transform.

The *finite Radon transform* (FRAT) is defined as summations of image pixels over a certain set of "lines". Those lines are defined in a finite geometry in a similar way as the lines for the continuous Radon transform in the Euclidean geometry. Denote $Z_p = \{0, 1, \dots, p-1\}$, where p is a prime number. Note that Z_p is a finite field with modulo p operations [23]. For later convenience, we denote $Z_p^* = \{0, 1, \dots, p\}$.

The FRAT of a real function f on the finite grid Z_p^2 is defined as

$$r_k[l] = FRAT_f(k, l) = \frac{1}{\sqrt{p}} \sum_{(i,j) \in L_{k,l}} f[i, j]. \qquad (8)$$

Here $L_{k,l}$ denotes the set of points that make up a line on the lattice Z_p^2, or more precisely

$$\begin{aligned} L_{k,l} &= \{(i, j) \;:\; j = ki + l \pmod p,\; i \in Z_p\}, \quad 0 \le k < p, \\ L_{p,l} &= \{(l, j) \;:\; j \in Z_p\}. \end{aligned} \qquad (9)$$

Figure 3 shows an example of the finite lines $L_{k,l}$ where points in the grid Z_p^2 are represented by image pixels. Note that due to the modulo operations in the definition of lines for the FRAT, these lines exhibit a "wrap around" effect. In other words, the FRAT treat the input image as one period of a periodic image. Later, we will present several ways to limit this artifact.

We observe that in the FRAT domain, the energy is best compacted if the mean is subtracted from the image $f[i, j]$ prior to taking the transform given in (8), which is assumed in the sequel. We also introduce the factor $p^{-1/2}$ in order to normalize the l_2-norm between the input and output of the FRAT.

Just as in the Euclidean geometry, a line $L_{k,l}$ on the affine plane Z_p^2 is uniquely represented by its slope or direction $k \in Z_p^*$ ($k = p$ corresponds to infinite slope or vertical lines) and its intercept $l \in Z_p$. One can verify that there are $p^2 + p$ lines defined in this way and every line contains p points. Moreover, any two distinct points on Z_p^2 belong to just one line. Also, two lines of different slopes intersect at exactly one point. For any given slope, there are p parallel lines that provide a complete cover of the plane Z_p^2. This means that for an input image $f[i, j]$ with zero-mean, we have

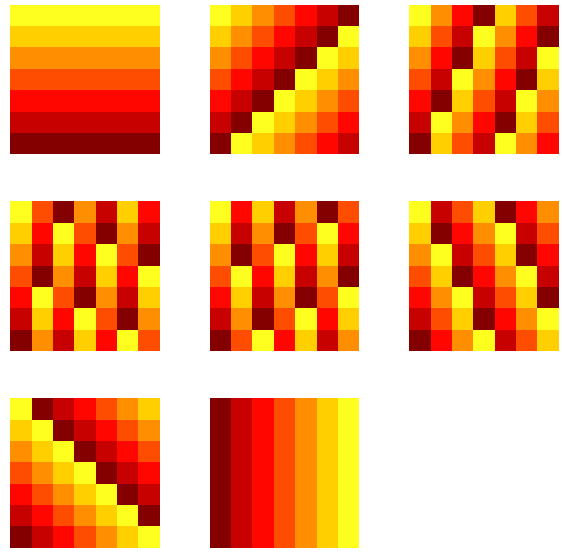


Fig. 3. Lines for the 7×7 FRAT. Parallel lines are grouped in each of the eight possible directions. Images in order from top to bottom, left to right are corresponding to the values of k from 0 to 7. In each image, points (or pixels) in different lines are assigned with different gray-scales.

$$\sum_{l=0}^{p-1} r_k[l] = \frac{1}{\sqrt{p}} \sum_{(i,j) \in Z_p^2} f[i, j] = 0 \quad \forall k \in Z_p^*. \qquad (10)$$

Thus, (10) explicitly reveals the redundancy of the FRAT: in each direction, there are only $p-1$ independent FRAT coefficients. Those coefficients at $p+1$ directions together with the mean value make up totally of $(p+1)(p-1)+1 = p^2$ independent coefficients (or degrees of freedom) in the finite Radon domain, as expected.

By analogy with the continuous case, the *finite back-projection* (FBP) operator is defined as the sum of Radon coefficients of all the lines that go through a given point, that is

$$FBP_r(i, j) = \frac{1}{\sqrt{p}} \sum_{(k,l) \in P_{i,j}} r_k[l], \quad (i, j) \in Z_p^2, \qquad (11)$$

where $P_{i,j}$ denotes the set of indices of all the lines that go through a point $(i, j) \in Z_p^2$. More specifically, using (9) we can write

$$P_{i,j} = \{(k, l) \;:\; l = j - ki \pmod p,\; k \in Z_p\} \cup \{(p, i)\}. \qquad (12)$$

From the property of the finite geometry Z_p^2 that every two points lie on exactly one line, it follows that every point in Z_p^2 lies on exactly one line from the set $P_{i,j}$, except for the point (i, j) which lies on all $p+1$ lines. Thus, by substituting (8) into (11) we obtain

$$\begin{aligned} FBP_r(i,j) &= \frac{1}{p}\sum_{(k,l)\in P_{i,j}}\sum_{(i',j')\in L_{k,l}} f[i',j'] \\ &= \frac{1}{p}\left(\sum_{(i',j')\in Z_p^2} f[i',j'] + p.f[i,j]\right) \\ &= f[i,j]. \end{aligned} \tag{13}$$

So the back-projection operator defined in (11) indeed computes the inverse FRAT for zero-mean images. Therefore we have an efficient and exact reconstruction algorithm for the FRAT. Furthermore, since the FBP operator is the adjoint of the FRAT operator, the algorithm for the inverse of FRAT has the same structure and is symmetric with the algorithm for the forward transform.

It is easy to see that the FRAT requires *exactly* p^3 additions and p^2 multiplications. Moreover, for memory access efficiency, [20] describes an algorithm for the FRAT in which for each projection k we need to pass through every pixel of the original image only once using p histogrammers, one for each summation in (8) of that projection. For images of moderate sizes, we observed that the actual computational time of the FRAT is compatible with other $O(p^2 \log(p^2))$ transforms, such as the 2-D FFT, where the leading constant can be large. For example, on a Sun Ultra 5 computer, both the forward and inverse FRAT's take less than a second to compute on an image of size 257×257.

B. Optimal Ordering of the Finite Radon Transform Coefficients

The FRAT described in Section III-A uses (9) as a convenient way of specifying finite lines on the Z_p^2 grid via two parameters: the slope k and the intercept l. However it is neither a unique nor the best way for our purpose. Let us consider a more general definition of lines on the finite Z_p^2 plane as

$$L'_{a,b,t} = \left\{(i,j)\in Z_p^2 \;:\; ai+bj-t = 0 \pmod p\right\}, \tag{14}$$

where $a, b, t \in Z_p$ and $(a,b) \neq (0,0)$.

This is by analogy with the line equation: $x_1\cos\theta + x_2\sin\theta - t = 0$ in $\mathbb{R}^2$. Therefore, for a finite line defined as in (14), (a,b) has the role of the normal vector, while t is the translation parameter. In this section, all equations involving line parameters are carried out in the finite field Z_p, which is assumed in the sequel without the indication of mod p.

It is easy to verify that for a fixed normal vector (a,b), $\left\{L'_{a,b,t} : t \in Z_p\right\}$ is a set of p parallel lines in the Z_p^2 plane. This set is equal to the set of p lines $\{L_{k,l} : l \in Z_p\}$ defined in (9) with the same slope k, where $k = -b^{-1}a$ for $b \neq 0$ and $k = p$ for $b = 0$. Moreover, the set of lines with the normal vector (a,b) is equal to the set of lines with the normal vector (na, nb), for each $n = 1, 2, \ldots, p-1$.

With the general line specification in (14), we now define the new FRAT to be

$$r_{a,b}[t] = FRAT_f(a,b,t) = \frac{1}{\sqrt{p}}\sum_{(i,j)\in L'_{a,b,t}} f[i,j]. \tag{15}$$

From the discussion above we see that a new FRAT projection sequence: $(r_{a,b}[0], r_{a,b}[1], \ldots, r_{a,b}[p-1])$, is simply a reordering of a projection sequence $(r_k[0], r_k[1], \ldots, r_k[p-1])$ from (8). This ordering is important for us since we later apply a 1-D wavelet transform on each FRAT projection. Clearly, the chosen normal vectors (a,b) control the order for the coefficients in each FRAT's projection, as well as the represented directions of those projections.

The usual FRAT described in Section III-A uses the set of $(p+1)$ normal vectors $\boldsymbol{u}_k$, where

$$\begin{aligned} \boldsymbol{u}_k &= (-k, 1) \quad \text{for } k = 0, 1, \ldots, p-1, \text{ and} \\ \boldsymbol{u}_p &= (1, 0). \end{aligned} \tag{16}$$

In order to provide a complete representation, we need the FRAT to be defined as in (15) with a set of $p+1$ normal vectors $\left\{(a_k, b_k) : k \in Z_p^*\right\}$ such that they cover all $p+1$ distinct FRAT projections represented by $\left\{\boldsymbol{u}_k : k \in Z_p^*\right\}$. We have $p-1$ choices for each of those normal vectors as

$$(a_k, b_k) = n\boldsymbol{u}_k, \quad 1 \le n \le p-1.$$

So what is the good choice for the $p+1$ normal vectors of the FRAT? To answer this we first prove the following projection slice theorem for the general FRAT. A special case of this theorem is already shown in [20].

Defining $W_p = e^{-2\sqrt{-1}\pi/p}$, the *discrete Fourier transform* (DFT) of a function f on Z_p^2 can be written as

$$F[u,v] = \frac{1}{p}\sum_{(i,j)\in Z_p^2} f[i,j] W_p^{ui+vj}, \tag{17}$$

and for FRAT projections on Z_p as

$$R_{a,b}[w] = \frac{1}{\sqrt{p}}\sum_{t\in Z_p} r_{a,b}[t] W_p^{wt}. \tag{18}$$

Theorem 1 (Discrete projection-slice theorem) The 1-D DFT $R_{a,b}[w]$ of a FRAT projection $r_{a,b}[t]$ is identical to the 2-D DFT $F[u,v]$ of $f[i,j]$ evaluated along a discrete slice through the origin at direction (a,b):

$$R_{a,b}[w] = F[aw, bw]. \tag{19}$$

Proof: Substituting (15) into (18) and using the fact that the set of p parallel lines $\left\{L'_{a,b,t} : t \in Z_p\right\}$ provides a complete cover of the plane Z_p^2, we obtain

$$\begin{aligned} R_{a,b}[w] &= \frac{1}{p}\sum_{t\in Z_p}\sum_{(i,j)\in L'_{a,b,t}} f[i,j] W_p^{wt} \\ &= \frac{1}{p}\sum_{(i,j)\in Z_p^2} f[i,j] W_p^{w(ai+bj)} \\ &= F[aw, bw]. \end{aligned}$$

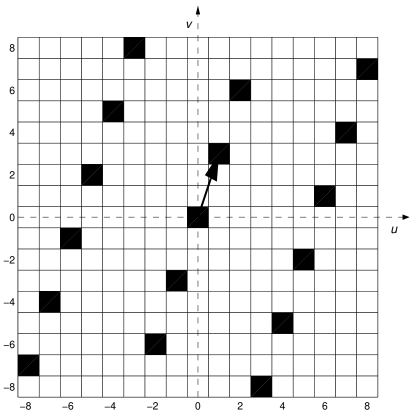


Fig. 4. Example of a discrete Fourier slice (indicated by the black squares) with the best normal vector for that FRAT projection. In this example, $p = 17$ and the slope $k = 11$. The normal vector can be chosen as a vector from the origin to any other points on the Fourier slide. The best normal vector is $(1, 3)$ (the solid arrow).

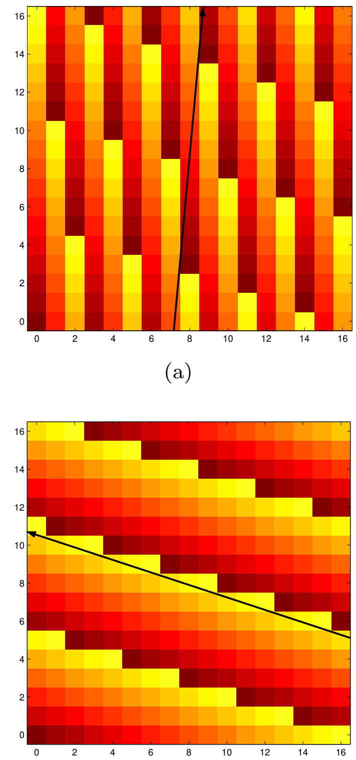
From (19), we can see the role of the FRAT normal vectors (a, b) in the DFT domain: it controls the order of the coefficients in the corresponding Fourier slices. In particular, $F[a, b]$ equals to the first harmonic component of the FRAT projection sequence with the normal vector (a, b). For the type of images that we are interested in, e.g. of natural scenes, most of the energy is concentrated in the low frequencies. Therefore in these cases, in order to ensure that each FRAT projection is smooth or low frequency dominated so that it can be represented well by the wavelet transform, the represented normal vector (a, b) should be chosen to be as "close" to the origin of the Fourier plane as possible.

Figure 4 illustrates this by showing an example of a discrete Fourier slice. The normal vector for the corresponding FRAT projection can be chosen as a vector from the origin to *any* other point on the Fourier slice. However, the best normal vector is selected as the closest point to the origin. The choice of the normal vector (a, b) as the closest point to the origin causes the represented direction of the FRAT projection to have the least "wrap around" due to the periodization of the transform. The effect of the new ordering of FRAT coefficient in the image domain is illustrated in Figure 5 for the same example projection. As can be seen, the "wrap around" effect is significantly reduced with the optimal ordering compared to the usual one.

Formally, we define the set of $p+1$ optimal normal vectors $\left\{(a_k^\star, b_k^\star) : k \in Z_p^*\right\}$ as follows

$$(a_k^\star, b_k^\star) = \arg \min_{\substack{(a_k, b_k) \in \{n\mathbf{u}_k : 1 \leq n \leq p-1\} \\ \text{s.t. } \mathcal{C}_p(b_k) \geq 0}} \|(\mathcal{C}_p(a_k), \mathcal{C}_p(b_k))\|. \tag{20}$$

Here $\mathcal{C}_p(x)$ denotes the centralized function of period p: $\mathcal{C}_p(x) = x - p.\text{round}(x/p)$. Hence, $\|(\mathcal{C}_p(a_k), \mathcal{C}_p(b_k))\|$ represents the distance from the origin to the point (a_k, b_k) on the periodic Fourier plane as shown in Figure 4. The constraint $\mathcal{C}_p(b_k) \geq 0$ is imposed in order to remove the ambiguity in deciding between (a, b) and $(-a, -b)$ as the normal vector for a projection. As a result, the optimal normal vectors are restricted to have angles in $[0, \pi)$. We use norm-2 for solving (20). Minimization is simply done for each $k \in Z_p^*$ by computing $p - 1$ distances in (20) and select the smallest one. Figure 6 shows an example of the optimal set of normal vectors. In comparison with the usual set of normal vectors $\left\{\boldsymbol{u}_k : k \in Z_p^*\right\}$ as given in (16), the new set $\left\{(a_k^\star, b_k^\star) : k \in Z_p^*\right\}$ provides a much more uniform angular coverage.



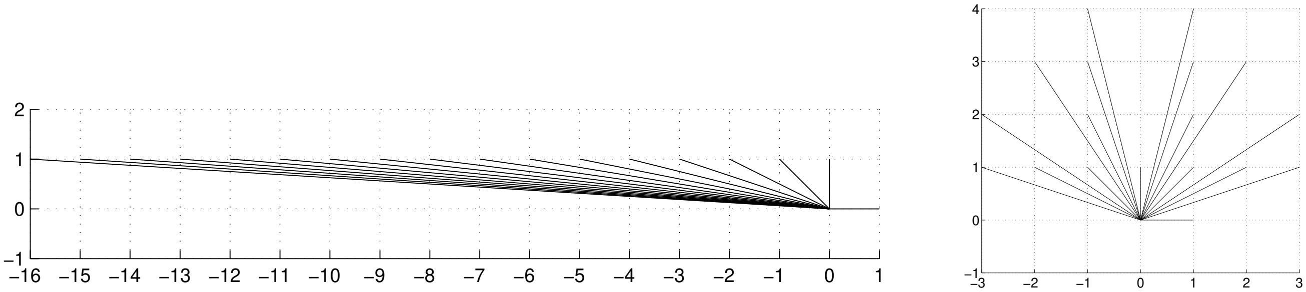
(a)

(b)

Fig. 5. Lines for the FRAT projection as shown in Figure 4 using: (a) usual ordering, (b) optimal ordering. They both represent the same set of lines but with different orderings. The orderings are signified by the increasing of gray-scales. The arrows indicate the represented directions in each case.

After obtaining the set of normal vectors $\{(a_k^\star, b_k^\star)\}$, we can compute the FRAT and its inverse with the same fast algorithms using histogrammers described in Section III-A. For a given p, solving (20) requires $O(p^2)$ operations and therefore it is negligible compared to the transforms themselves. Furthermore, this can be pre-computed, thus only presents as a "one-time" cost.

For the sake of simplicity, we write $r_k[t]$ for $r_{a_k^\star, b_k^\star}[t]$ in the sequel. In other words, from now we regard k as an index in the set of optimal FRAT normal vectors rather than a slope. Likewise, the line $L'_{a_k^\star, b_k^\star, t}$ is simply rewritten as $L_{k,t}$, for $0 \leq k \leq p,\ 0 \leq t < p$.



(a) (b)

Fig. 6. The set of normal vectors, which indicate the represented directions, for the FRAT of size $p = 17$ using: (a) usual ordering; (b) optimal ordering.

C. Frame Analysis of the FRAT

Since the FRAT is a redundant transform, it can be studied as a *frame operator*. In this section we will study the FRAT in more detail and reveal some of its properties in this frame setting. A detailed introduction to frames can be found in [24], [3].

Suppose that $\mathcal{F}$ is a linear operator from $\mathbb{R}^N$ to $\mathbb{R}^M$, defined by

$$(\mathcal{F}x)_n = \langle x, \varphi_n \rangle, \qquad \text{for } n = 1, \ldots, M. \tag{21}$$

The set $\{\varphi_n\}_{n=1}^{M} \subset \mathbb{R}^N$ is called a frame of $\mathbb{R}^N$ if there exist two constants $A > 0$ and $B < \infty$ such that

$$A\|x\|^2 \leq \sum_{n=1}^{M} |\langle x, \varphi_n \rangle|^2 \leq B\|x\|^2, \qquad \forall x \in \mathbb{R}^N. \tag{22}$$

where A and B are called the frame bounds. When $A = B$ the frame is said to be tight. If the frame condition is satisfied then $\mathcal{F}$ is called a frame operator. It can be shown that any finite set of vectors that spans $\mathbb{R}^N$ is a frame. The frame bound ratio B/A indicates the numerical stability in reconstructing x from $(\mathcal{F}x)_n$; the tighter the frame, the more stable the reconstruction against coefficient noise.

The frame operator can be regarded as a left matrix multiplication with F, where F is an $M \times N$ matrix in which its nth row equals to φ_n. The frame condition (22) can be rewritten as

$$x^T A x \leq x^T F^T F x \leq x^T B x, \qquad \forall x \in \mathbb{R}^N. \tag{23}$$

Since $F^T F$ is symmetric, it is diagonalizable in an orthonormal basis [25], thus (23) implies that the eigenvalues of $F^T F$ are between A and B. Therefore, the tightest possible frame bounds A and B are the minimum and maximum eigenvalues of $F^T F$, respectively. In particular, a tight frame is equivalent to $F^T F = A \cdot I$, which means the transpose of F equals to its left inverse within a scale factor A.

Now let us return to the FRAT. Since it is invertible it can be regarded as a frame operator in $l_2(Z_p^2)$ with the frame $\left\{\varphi_{k,l} : k \in Z_p^*, l \in Z_p\right\}$ defined as

$$\varphi_{k,l} = p^{-1/2} \delta_{L_{k,l}} \tag{24}$$

where δ_S denotes the characteristic function for the set S, which means $\delta_S[i,j]$ equals to 1 if $(i,j) \in S$ and 0 otherwise. Note that this frame is normalized since $\|\varphi_{(k,l)}\| = 1$. By writing images as column vectors, the FRAT can be regarded as a left matrix multiplication with $F = p^{-1/2}R$, where $\{R\}_{(k,l),\ (i,j)}$ is the $(p^2 + p) \times p^2$ incidence matrix of the affine geometry Z_p^2: $R_{(k,l),\ (i,j)}$ equals to 1 if $(i,j) \in L_{k,l}$ and 0 otherwise.

Proposition 1: The tightest bounds for the FRAT frame $\left\{\varphi_{k,l} : k \in Z_p^*, l \in Z_p\right\}$ in $l_2(Z_p^2)$ are $A = 1$ and $B = p + 1$.

Proof: From (23), these tightest bounds can be computed from the eigenvalues of $C = F^T F = p^{-1} R^T R$. Since R is the incidence matrix for lines in Z_p^2, $(R^T R)_{(i,j),\ (i',j')}$ equals the number of lines that go through both (i,j) and (i',j'). Using the properties of the finite geometry Z_p^2 that every two points lie in exactly one line and that there are exactly $p+1$ lines that go through each point, it follows that the entries of C equal to $(p+1)p^{-1}$ along its diagonal and p^{-1} elsewhere.

The key observation is that C is a circulant matrix, hence its eigenvalues can be computed as the p^2-points discrete Fourier transform (DFT) on its first column $c = ((p+1)p^{-1}, p^{-1}, \ldots, p^{-1})$ [1] (§2.4.8). Writing c as

$$c = (1, 0, \ldots, 0) + p^{-1} \cdot (1, 1, \ldots, 1),$$

we obtain,

$$DFT\{c\} = (1, 1, \ldots, 1) + p \cdot (1, 0, 0, \ldots, 0) = (p+1, 1, 1, \ldots, 1)$$

where the DFT is computed for the Dirac and constant signals.

Therefore the eigenvalues of C are $p+1$ and 1, the latter with multiplicity of $p^2 - 1$. As a result, the tightest (normalized) frame bounds for FRAT as $A = 1$ and $B = p + 1$. ∎

For reconstruction, the FBP defined in (11) can be represented by a left multiplication with matrix $p^{-1/2}B$, where $B_{(i,j),\ (k,l)}$ equals to 1 if $(k,l) \in P_{i,j}$ and 0 otherwise. From the definition of $P_{i,j}$, we have

$$R_{(k,l),\ (i,j)} = B_{(i,j),\ (k,l)}, \quad \forall i, j, k, l.$$

So the transform matrices for the operators FRAT and FBP are transposed of each other. Let $\bar{Z}_p^2$ denotes the subspace of zero-mean images defined on Z_p^2. Since the FBP is an inverse of the FRAT for zero-mean images, we have the following result.

Proposition 2: On the subspace of zero-mean images $\bar{Z}_p^2$, the FRAT is a *normalized tight frame* with $A = B = 1$, which means

$$f = \sum_{k=0}^{p} \sum_{l=0}^{p-1} \langle f, \varphi_{k,l} \rangle \varphi_{k,l}, \quad \forall f \in \bar{Z}_p^2. \tag{25}$$

Remark 1: It is instructive to note that constant images on Z_p^2 are eigenvectors of $C = F^T F$ with the eigenvalue $p + 1$. Taking constant images out leaves a system with all unity eigenvalues, or a tight frame on the remaining subspace. Thus, we have another interpretation of FRAT being a normalized tight frame for zero-mean images.

By subtracting the mean from the image before applying the FRAT, we change the frame bound ratio from $p+1$ to 1 and obtain a tight frame. Consequently, this makes the reconstruction more robust against noise on the FRAT coefficients due to thresholding and/or quantization. This follows from the result in [26] that with the additive white noise model for the coefficients, the tight frame is optimal among normalized frames in minimizing mean-squared error.

IV. Orthonormal Finite Ridgelet Transform

With an invertible FRAT and applying (6), we can obtain an invertible discrete ridgelet transform by taking the discrete wavelet transform (DWT) on each FRAT projection sequence, $(r_k[0], r_k[1], \ldots, r_k[p-1])$, where the direction k is fixed. We call the overall result the *finite ridgelet transform* (FRIT). Figure 7 depicts these steps.

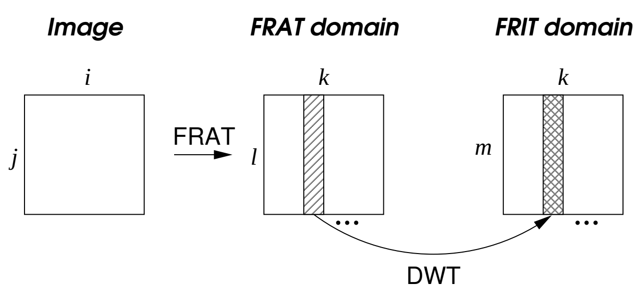


Fig. 7. Diagram for the FRIT. After taking the FRAT, a DWT is applied on each of the FRAT slices or projections where k is fixed.

Typically p is not dyadic, therefore a special border handling is required. Appendix A details one possible way of computing the DWT for prime length signals. Due to the periodicity property of the FRAT coefficients for each direction, periodic wavelet transforms are chosen and assumed in this section.

Recall that the FRAT is redundant and not orthogonal. Next we will show that by taking the 1-D DWT on the projections of the FRAT in a special way, we can remove this redundancy and obtain an orthonormal transform.

Assume that the DWT is implemented by an orthogonal tree-structured filter bank with J levels, where G_0 and G_1 are low and high pass synthesis filters, respectively. Then the family of functions:

$$\left\{ g_0^{(J)}[\cdot - 2^J m],\ g_1^{(j)}[\cdot - 2^j m] \ : \ j = 1, \ldots, J;\ m \in \mathbb{Z} \right\}$$

is the orthogonal basis of the discrete-time wavelet series [1]. Here $G^{(j)}$ denotes the equivalent synthesis filters at level j, or more specifically

$$\begin{aligned} G_0^{(J)}(z) &= \prod_{k=0}^{J-1} G_0(z^{2^k}), \\ G_1^{(j)}(z) &= G_1(z^{2^{j-1}}) \prod_{k=0}^{j-2} G_0(z^{2^k}), \quad j = 1, \ldots, J. \end{aligned}$$

The basis functions from $G_0^{(J)}$ are called the scaling functions, while all the others functions in the wavelet basis are called wavelet functions. Typically, the filter G_1 is designed to satisfy the high pass condition, $G_1(z)|_{z=1} = 0$ so that the corresponding wavelet has at least one vanishing moment. Therefore, $G_1^{(j)}(z)|_{z=1} = 0$, $\forall j = 1, \ldots, J$, which means all wavelet basis functions have zero mean.

For a more general setting, let us assume that we have a collection of $(p + 1)$ 1-D orthonormal transforms on $\mathbb{R}^p$ (which can be the same), one for each projection k of FRAT, that have bases as

$$\left\{ \boldsymbol{w}_m^{(k)} \ : \ m \in Z_p \right\}, \quad k = 0, 1, \ldots, p.$$

The only condition that we require for each of these bases can be expressed equivalently by the following lemma.

Lemma 1 (Condition Z) Suppose that $\{\boldsymbol{w}_m : m \in Z_p\}$ is an orthogonal basis for the finite-dimensional space $\mathbb{R}^p$, then the following are equivalent:

1. This basis contains a constant function, say $\boldsymbol{w}_0$, i.e. $w_0[l] = const$, $\forall l \in Z_p$.
2. All other basis functions, $\boldsymbol{w}_m$, $m = 1, \ldots, p-1$, have zero mean.

Proof: Denote $\mathbf{1} = (1, 1, \ldots, 1) \in \mathbb{R}^p$. If $\boldsymbol{w}_0 = c\mathbf{1}$, $c \neq 0$ then from the orthogonality assumption that $\langle \boldsymbol{w}_0, \boldsymbol{w}_m \rangle = 0$, we obtain $\sum_l w_m[l] = 0$, $\forall m = 1, \ldots, p-1$.

Conversely, assume that each basis function $\boldsymbol{w}_m$, $1 \leq m \leq p-1$, has zero mean. Denote S the subspace that is spanned by these functions and $S^\perp$ is its orthogonal complement subspace in $\mathbb{R}^p$. It is clear that $S^\perp$ has dimension 1 with $\boldsymbol{w}_0$ as its basis. Consider the subspace $S_0 = \{c\mathbf{1} : c \in \mathbb{R}\}$. We have $\langle c\mathbf{1}, \boldsymbol{w}_m \rangle = c \sum_l w_m[l] = 0$, $\forall m = 1, \ldots, p-1$, thus $S_0 \subset S^\perp$. On the other hand, $\dim(S_0) = \dim(S^\perp) = 1$, therefore $S^\perp = S_0$. This means $\boldsymbol{w}_0$ is a constant function. ∎

As shown before, the *Condition Z* is satisfied for all wavelet bases, or in fact any general tree-structured filter banks where the all-lowpass branch is carried to the maximum number of stages (i.e. when only one scaling coefficient is left).

By definition, the FRIT can be written as

$$\begin{aligned} FRIT_f[k,m] &= \langle FRAT_f[k,\cdot], w_m^{(k)}[\cdot]\rangle \\ &= \sum_{l\in Z_p} w_m^{(k)}[l]\langle f, \varphi_{k,l}\rangle \\ &= \langle f, \sum_{l\in Z_p} w_m^{(k)}[l]\ \varphi_{k,l}\rangle. \end{aligned} \tag{26}$$

Here $\{\varphi_{k,l}\}$ is the FRAT frame which is defined in (24). Hence we can write the basis functions for the FRIT as follows:

$$\rho_{k,m} = \sum_{l\in Z_p} w_m^{(k)}[l]\ \varphi_{k,l}. \tag{27}$$

We can next prove the result on the orthogonality of a modified FRIT.

Theorem 2: Given $p+1$ orthonormal bases in $l^2(Z_p)$ (which can be the same): $\left\{\boldsymbol{w}_m^{(k)} : m \in Z_p\right\},\ 0 \le k \le p$, that satisfy the *Condition Z* then

$$\{\rho_{k,m}\ :\ k=0,1,\ldots,p;\ m=1,2,\ldots,p-1\} \cup \{\rho_0\}$$

is an orthonormal basis in $l^2(Z_p^2)$, where $\rho_{k,m}$ are defined in (27) and ρ_0 is the constant function, $\rho_0[i,j]=1/p,\ \forall (i,j)\in Z_p^2$.

Proof: Let us consider the inner products between any two FRIT basis functions

$$\langle \rho_{k,m}, \rho_{k',m'}\rangle = \sum_{l,l'\in Z_p} w_m^{(k)}[l]\ w_{m'}^{(k')}[l']\ \langle \varphi_{k,l}, \varphi_{k',l'}\rangle.$$

Using properties of lines in the finite geometry Z_p^2, it is easy to verify that

$$\langle \varphi_{k,l}, \varphi_{k',l'}\rangle = \begin{cases} 1 & \text{if } k=k', l=l' \\ 0 & \text{if } k=k', l\neq l' \\ 1/p & \text{if } k\neq k' \end{cases} \tag{28}$$

Thus, when the two FRIT basis functions have the same direction, $k=k'$, then

$$\langle \rho_{k,m}, \rho_{k,m'}\rangle = \sum_{l\in Z_p} w_m^{(k)}[l]\ w_{m'}^{(k)}[l] = \delta[m-m'].$$

So the orthogonality of these FRIT basis functions comes from the orthogonality of the basis $\left\{\boldsymbol{w}^{(k)} : m\in Z_p\right\}$. In particular, we see that $\rho_{k,m}$ have unit norm. Next, for the case when the two FRIT basis functions have different directions, $k\neq k'$, using (28) we obtain

$$\begin{aligned} \langle \rho_{k,m}, \rho_{k',m'}\rangle &= \frac{1}{p}\sum_{l,l'\in Z_p} w_m^{(k)}[l]\ w_{m'}^{(k')}[l'] \\ &= \frac{1}{p}\left(\sum_{l\in Z_p} w_m^{(k)}[l]\right)\left(\sum_{l'\in Z_p} w_{m'}^{(k')}[l']\right). \end{aligned}$$

In this case, if either m or m' is non-zero, e.g. $m\neq 0$, then using the *Condition Z* of these bases, $\sum_{l\in Z_p} w_m^{(k)}[l] = 0$, it implies $\langle \rho_{k,m}, \rho_{k',m'}\rangle = 0$.

Finally, note that $\bigcup_l L_k(l) = Z_p^2$, for all directions k (see (10)). So, together with the assumption that $\boldsymbol{w}_0^{(k)}$ are constant functions, we see that all of the FRIT basis functions $\rho_{k,0},\ (k=0,1,\ldots,p)$ correspond to the mean of the input image so we only need to keep one of them (in any direction), which is denoted as ρ_0. The proof is now complete. ■

Remark 2: 1. An intuition behind the above result is that at each level of the DWT decomposition applied on the FRAT projections, all of the non-orthogonality and redundancy of the FRAT is pushed into the scaling coefficients. When the DWT's are taken to the maximum number of levels then all of the remaining scaling coefficients at different projections are the same, hence we can drop all but one of them. The result is an orthonormal FRIT.

2. We prove the above result for the general setting where different transforms can be applied on different FRAT projections. The choice of transforms can be either adaptive, depending on the image, or pre-defined. For example, one could employ an adaptive wavelet packet scheme independently on each projection. The orthogonality holds as long as the "all lowpass" branch of the general tree-structured filter bank is decomposed to a single coefficient. All other branches would contain at least one highpass filter thus leading to zero-mean basis functions.

3. Furthermore, due to the "wrap around" effect of the FRAT, some of its projections could contain strong periodic components so that a more oscillated basis like the DCT might be more efficient. Also note that from Theorem 1, if we apply the 1-D Fourier transform on all of the FRAT projections then we obtain the 2-D Fourier transform. For convenience, we still use the term FRIT to refer to the cases where other transforms than the DWT might be applied to some of the FRAT projections.

To gain more insight into the construction for the orthogonal FRIT basis, Figure 8 illustrates a simple example of the transform on a 2×2 block using the Haar wavelet. In this case, the FRIT basis is the same as the 2-D Haar wavelet basis, as well as the 2-D discrete Fourier basis.

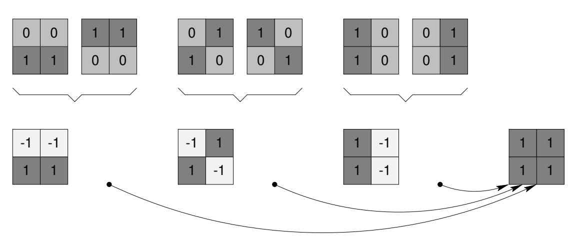


Fig. 8. Illustration on the contraction of orthogonal FRIT basis for a 2×2 block using the Haar wavelet. *Upper:* Basis images for the FRAT. *Lower:* Basis images for the orthogonal FRIT. These images are obtained by taking the (scaled) Haar transform for each pair (corresponding to one projection) of the FRAT basis images. The constant image results from all projections and thus we can drop all but one of them.

V. Variations on the Theme

A. Folded FRAT and FRIT

The FRAT in the previous sections is defined with a periodic basis over Z_p^2. This is equivalent to applying the transform to a periodization of the input image f. Therefore relatively large amplitude FRAT coefficients could result due to the possible discontinuities across the image borders. To overcome this problem, we propose a similar strategy as in the block cosine transform by extending the image symmetrically about its borders [3].

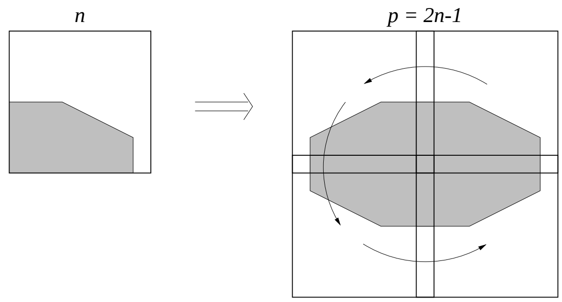


Fig. 9. Extending the image symmetrically about its borders in order to reduce the discontinuities across the image borders due to the periodization.

Given that p is a prime number and $p > 2$, then p is odd and can be written as $p = 2n-1$. Consider an $n \times n$ input image $f[i,j]$, $0 \le i,j < n$. Fold this image with respect to the lines $i = 0$ and $j = 0$ to produce a $p \times p$ image $\check{f}[i,j]$, in which (also see Figure 9)

$$\check{f}[i,j] = f[|i|,|j|], \quad -n < i,j < n. \tag{29}$$

The periodization of $\check{f}[i,j]$ is symmetric and continuous across the borders of the original image, thus eliminating the jump discontinuity that would have resulted from the periodic extension of $f[i,j]$. Applying the FRAT to the $\check{f}[i,j]$ results in $p(p+1)$ transform coefficients. Notice the new range for the pixel indices of the image $\check{f}[i,j]$. We will show that the FRAT coefficients of $\check{f}[i,j]$ exhibit certain symmetry properties so that the original image can be perfectly reconstructed by keeping exactly n^2 coefficients.

Consider the 2-D DFT of $\check{f}[i,j]$

$$\check{F}[u,v] = \frac{1}{p} \sum_{-n<i,j<n} \check{f}[i,j] W_p^{ui+vj}.$$

Using the symmetry property of $\check{f}[i,j]$ in (29), we obtain

$$\check{F}[u,v] = \check{F}[|u|,|v|].$$

Theorem 1 shows that the FRAT $\check{r}_{a,b}[t]$, $(-n < t < n)$ of $\check{f}[i,j]$ can be computed from the inverse 1-D DFT as

$$\check{r}_{a,b}[t] = \frac{1}{\sqrt{p}} \sum_{-n<w<n} \check{R}_{a,b}[w] W_p^{-wt},$$

where $\check{R}_{a,b}[w] = \check{F}[aw, bw]$. The symmetry of $\check{F}[u,v]$ thus yields

$$\check{R}_{a,b}[w] = \check{R}_{a,b}[|w|] \quad \text{and} \tag{30}$$
$$\check{R}_{a,b}[w] = \check{R}_{|a|,|b|}[w]. \tag{31}$$

From (30) we have $\check{r}_{a,b}[t] = \check{r}_{a,b}[|t|]$ or each projection $\check{r}_{a,b}[t]$ is symmetric about $t = 0$, and (31) reveals the duplications among those projections. In fact, with the set of optimal normal vectors in (20), except for two projections indexed by $(1,0)$ and $(0,1)$ (the vertical and horizontal projections, respectively) all other projections have an identical twin. By removing those duplications we are left with $2+(p-2)/2 = n+1$ projections. For example, we can select the set of $n+1$ independent projections as the ones with normal vectors in the first quadrant (refer to Figure 6). Furthermore, as in (10), the redundancy among the projections of the folded FRAT can be written as

$$\check{r}_{a_k^\star,b_k^\star}[0] + 2\sum_{t=1}^{n-1} \check{r}_{a_k^\star,b_k^\star}[t] = \frac{1}{\sqrt{p}} \sum_{-n<i,j<n} \check{f}[i,j]. \tag{32}$$

The next proposition summarizes the above results.

Proposition 3: The image $f[i,j]$ can be perfectly reconstructed from the following n^2-1 coefficients:

$$\check{r}_{a_k^\star,b_k^\star}[t] \quad \text{such that } \mathcal{C}_p(a_k^\star) \ge 0 \text{ and } 0 < t < n, \tag{33}$$

and the mean of the image $\check{f}[i,j]$.

To gain better energy compaction, the mean should be subtracted from the image $\check{f}[i,j]$ previous to taking the FRAT. The set of independent coefficients in (33) is referred as the folded FRAT of the image $f[i,j]$.

However, orthogonality might be lost in the folded FRIT (resulting from applying 1-D DWT on $n+1$ projections of the folded FRAT), since the basis functions from a same direction of the folded FRAT could have overlap. Nevertheless, if we loosen up the orthogonality constraint, then by construction, the folded FRAT projections $(\check{r}_{a_k^\star,b_k^\star}[t] : 0 < t < n)$ are symmetric with respect to $t = 0$ and $t = n - 1/2$. This allows the use of folded wavelet transform with biorthogonal symmetric wavelets [27] or orthogonal symmetric IIR wavelets [28]. We anticipate the folded FRIT has potential in block transforms (i.e. dividing the image into small blocks and applying FRIT to each block) where the border effect is more serious, and plan report the results in a forthcoming paper.

B. Multilevel FRIT's

In the FRIT scheme described previously, multiscale comes from the 1-D DWT. As a result, at each scale, there is a large number of directions, which is about the size of the input image. Moreover, the basis images of the FRIT have long support, which extend over the whole image.

Here we propose a different scheme where the number of directions can be controlled, and the basis functions have smaller support. Assume that the input image has the size $n \times n$, where $n = p_1 p_2 \ldots p_J q$ and p_i are prime numbers. First, we apply the orthonormal FRIT to $n_1 \times n_1$ non-overlapping subimages of size $p_1 \times p_1$, where $n_1 = p_2 \ldots p_J q$. Each sub-image is transformed into $p_1^2 - 1$ "detail" FRIT coefficients plus a mean value. These mean values form an $n_1 \times n_1$ coarse approximate image of the original one. Then the process can be iterated on the coarse version up to J levels. The result is called as multilevel FRIT (MFRIT).

At each level, the basis functions for the "detail" MFRIT coefficients are obviously orthogonal within each block, and also with other blocks since they do not overlap. Furthermore, these basis functions are orthogonal with the constant function on their block, and thus orthogonality holds across levels as well. Consequently, the MFRIT is an orthonormal transform.

By collecting the MFRIT coefficients into groups depending on their scales and directions, we obtain a subband-like decomposition with J scales, where level i has p_i directions. When $p_i = 2$, the orthonormal FRIT using the Haar DWT is the same as the 2×2 Haar DWT (see Figure 8). Therefore the MFRIT scheme includes the multilevel 2-D Haar DWT. In general, when $p_i > 2$, the MFRIT offers more directions than the 2-D DWT and can be useful in certain applications such as texture analysis.

VI. Numerical Experiments

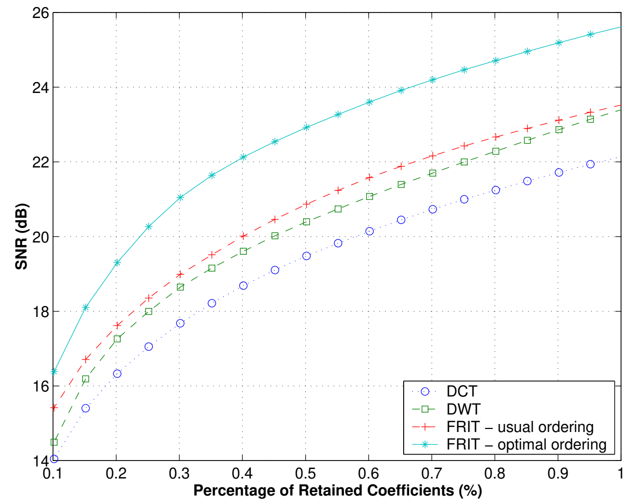
A. Non-linear Approximation

Following the study of the efficiency of the ridgelet transform in the continuous domain using the truncated Gaussian functions [6], we first perform numerical comparison on a 256×256 image of the function: $f(x_1, x_2) = 1_{\{x_2 < 2x_1 + 0.5\}} e^{-x_1^2 - x_2^2}$ (see Figure 10(a)), using four 2-D transforms: DCT, DWT, FRAT and FRIT. The comparison is evaluated in terms of the non-linear approximation power, i.e. the ability of reconstructing the original image, measured by signal-to-noise ratios (SNR's), using the N largest magnitude transform coefficients. For the FRAT and FRIT, we extend the image size to the next prime number, 257, by replicating the last pixel in each row and column. We use the orthogonal *Symmlet* wavelet with 4 vanishing moments [24] for both the DWT and the FRIT.

Our initial experiments indicate that in order to achieve good results, it is necessary to apply strong oscillated bases to certain FRAT projections to handle to the "wrap around" effect (refer to the remarks at the end of Section IV). For images with linear singularities, we find that in the FRAT domain, most of the image energy and singularities are contained in the projections with the *least* "wrap around" (see Figure 13(b)). Therefore, without resorting to adaptive methods, we employ a simple, predefined scheme where the DWT is only applied to the projections with $\|(a_k^\star, b_k^\star)\| \leq D$, while the remaining projections use the DCT. We use $D = 3$ in our experiments, which means in the tested FRIT, only 16 FRAT projections are represented by the DWT. Although the this FRIT contains most of Fourier-type basis functions, due to the concentration of energy mentioned above, the resulting nonlinear approximation images are mainly composed of the ridgelet-type functions that fit around the linear edge.

Figure 10(b) display the comparison results. We omit the FRAT since its performance is much worse than the others. Clearly the FRIT achieves the best result, as expected from the continuous theory. Furthermore, the new ordering of the FRAT coefficients is crucial for the FRIT in obtaining good performance.

(a)



(b)

Fig. 10. (a) Test image: a truncated Gaussian image of size 256×256 that represents the function $f(x_1, x_2) = 1_{\{x_2 < 2x_1 + 0.5\}} e^{-x_1^2 - x_2^2}$. (b) Comparison of non-linear approximations using four different 2-D transforms: DCT, DWT, FRIT with usual ordering and FRIT with optimal ordering.

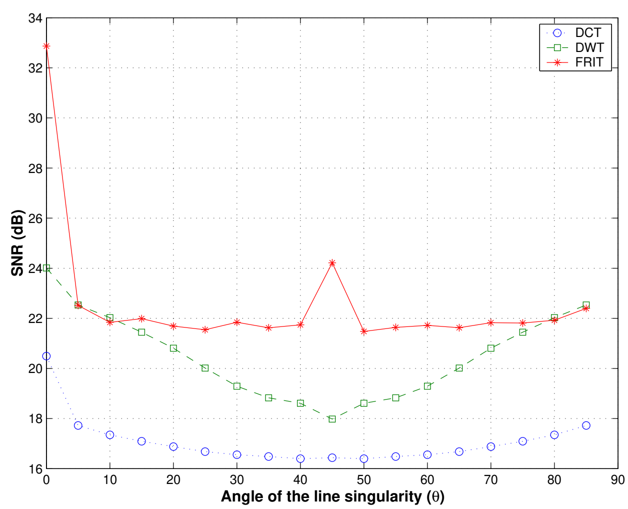
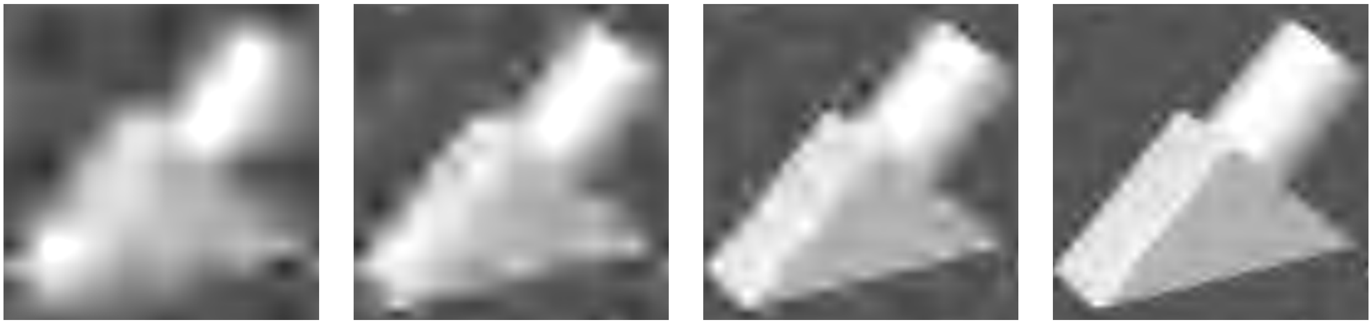
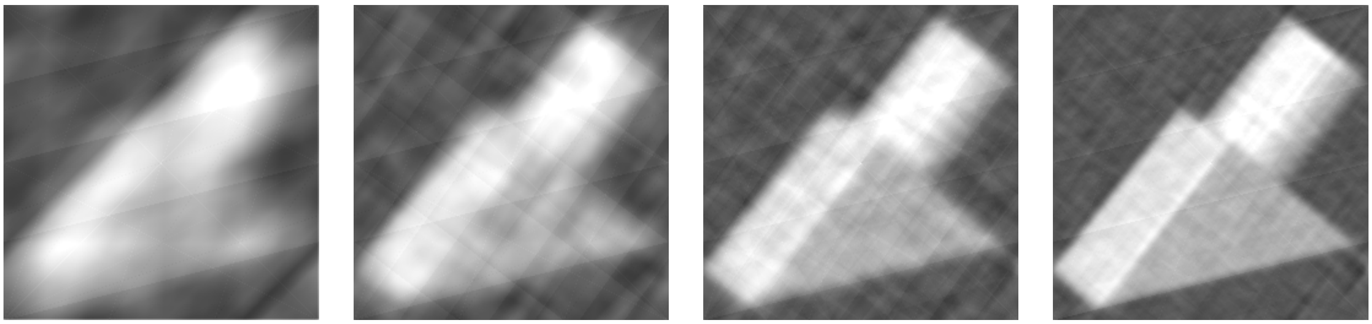


Fig. 11. Non-linear approximation comparison at different orientation of the line singularity in the truncated Gaussian images $f_\theta(x_1, x_2) = 1_{\{x_1 \cos\theta + x_2 \sin\theta < 0.3\}} e^{-x_1^2 - x_2^2}$. In each case, we keep the most 0.5% significant coefficients.



(a)Using DWT



(b)Using FRIT

Fig. 12. From left to right, reconstructed images from the 32, 64, 128 and 256 most significant coefficients of the DWT and FRIT, out of 65536 coefficients.

We then compare the performance where the singularity line varies its orientation. Consider the truncated Gaussian image again, using the function $f_\theta(x_1, x_2) = 1_{\{x_1 \cos\theta + x_2 \sin\theta < 0.3\}} e^{-x_1^2 - x_2^2}$. Due to the circular symmetry, we only need to consider $0 \leq \theta < 90°$. Figure 11 shows the results where the FRIT (with optimal ordering) consistently outperforms both the DWT, more than 2 dB on the average, as well as the DCT.

Our next test is a real image of size 256×256 with straight edges. Figure 12 shows the images obtained from non-linear approximation using the DWT and FRIT. As can be seen, the FRIT correctly picks up the edges using the first few significant coefficients and produces visually better approximated images. But let us point out that even this simple test image can not be represented as a summation of a few "global" linear singularities (like the Gaussian truncated images), and thus it is *not* in the optimal class of the ridgelet transform.

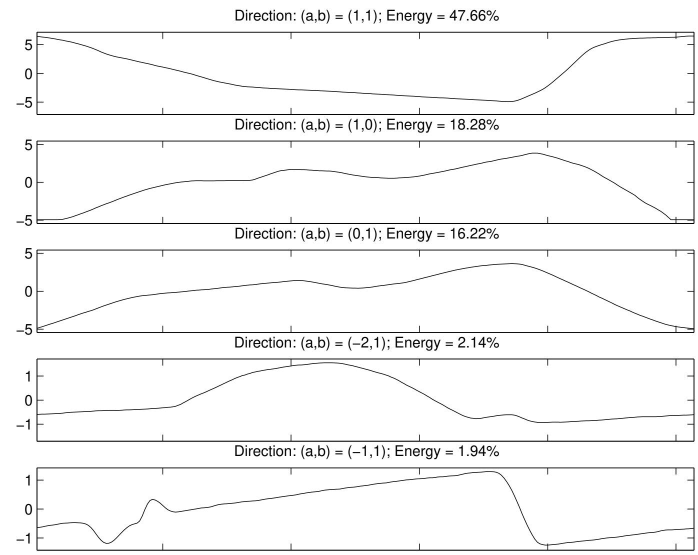
To gain an insight into the FRIT, Figure 13(a) shows the top five FRAT projections for the "object" image that contain most of the energy, measured in the l_2-norm. Those projections correspond to the directions that have discontinuities across, plus the horizontal and vertical directions. Therefore, we see that at first the FRAT compacts most of the energy of the image into a few projections (see Figure 13(b)), where the linear discontinuities create "jumps". Next, taking the 1-D DWT on those projections, which are mainly smooth, compacts the energy further into a few FRIT coefficients.

B. *Image Denoising*

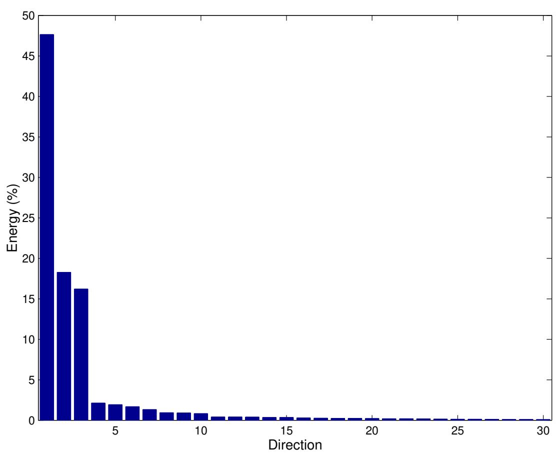
The motivation for the FRIT-based image denoising method is that in the FRIT domain, linear singularities of the image are represented by a few large coefficients, whereas randomly located noisy singularities are unlikely to produce significant coefficients. By contrast, in the DWT domain, both image edges and noisy pixels produce similar amplitude coefficients. Therefore, a simple thresholding scheme for FRIT coefficients can be very effective in denoising images that are piecewise smooth away from singularities along straight edges.

We consider a simple case where the original image is contaminated by an additive zero-mean Gaussian white noise of variance σ^2. With an orthogonal FRIT, the noise in the transform domain is also Gaussian white of the same variance. Therefore it is appropriate to apply the thresholding estimators that were proposed in [29] to the FRIT coefficients. More specifically, our denoising algorithm consists of the following steps:

Step 1: Applying FRIT to the noisy image.
Step 2: Hard-thresholding of FRIT coefficients with the universal threshold $T = \sigma\sqrt{2 logN}$ where $N = p^2$ pixels.
Step 3: Inverse FRIT of the thresholded coefficients.
For an image which is smooth away from linear singularities, edges are visually well restored after Step 3. However due to the periodic property of the FRIT, strong edges sometimes create "wrap around" effects which are visible in the smooth regions of the image. In order to overcome this problem, we optionally employ a 2-D adaptive filtering step.



(a)



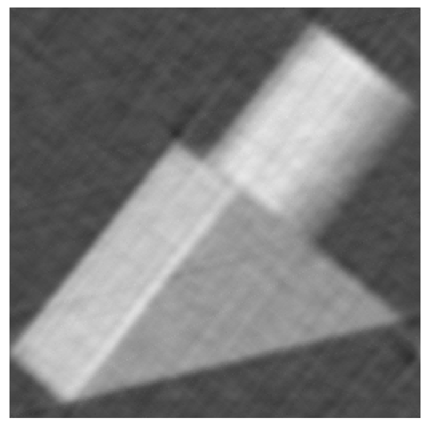
(b)

Fig. 13. (a) Top five FRAT projections of the "object" image that contain most of the energy. (b) Distribution of total input image energy among FRAT projections. Only the top 30 projections are shown in the descending order.

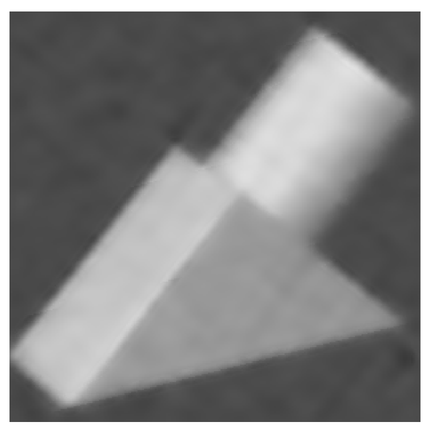
Step 4: (Optional) Adaptive Wiener filtering to reduce the "wrap around" effect.
In some cases, this can enhances the visual appearance of the restored image.

The above FRIT denoising algorithm is compared with the analogous wavelet hard-thresholding method using the same threshold value. Figure 14 shows the denoising results on the real image. The FRIT is clearly shown to be more effective than the DWT in recovering straight edges, as well as in term of SNR's.

(a)Using DWT; SNR = 19.78 dB.



(b)Using FRIT; SNR = 19.67 dB



(c)Using FRIT and Wiener filter; SNR = 21.07 dB.

Fig. 14. Comparison of denoising on the "object" image.

VII. Conclusion and Discussion

We presented a new family of discrete orthonormal transforms for images based on the ridgelet idea. Owning to orthonormality, the proposed ridgelet transform is self-inverting – the inverse transform uses the same algorithm as the forward transform – and has excellent numerical stability. Experimental results indicate that the FRIT offers an efficient representation for images that are smooth away from line discontinuities or straight edges. A Matlab code implementing the transforms and experiments in this paper is available at an author's Web page `www.ifp.uiuc.edu/~minhdo`.

However, it is important to emphasize that the ridgelet transform is only suited for discontinuities along straight lines. For complex images, where edges are mainly along curves and there are texture regions (which generate point discontinuities), the ridgelet transform is not optimal. Therefore, a more practical scheme in employing the ridgelet transform is to first utilize a quad-tree division of images into suitable blocks where edges look straight and then apply the discrete ridgelet transform to each block. Another scheme is to use the ridgelet transform as the building block in a more localized construction such as the *curvelet* transform [30].

Acknowledgment The authors thank Prof. Attila Kuba and Prof. Gabor Herman for pointing us the references [20], [21], Pier Luigi Dragotti and Rahul Shukla of EPFL for many stimulating discussions, and the reviewers for their constructive remarks.

Appendix

I. Orthogonal Wavelet Transform for Non Dyadic Length Signals

In the construction of the orthonormal FRIT, we need wavelet bases for signals of prime length p. In addition, those bases have to satisfy the *Condition Z* in Lemma 1. Let $n = 2^J$ be the nearest dyadic number to p that is smaller than or equal to p. Suppose that $p - n$ is small, then one simple way of taking the wavelet transform on a sequence of p samples is to apply the usual wavelet transform on the first n samples and then extend it to cover the remaining $p - n$ samples.

Let $\{\boldsymbol{v}_m : m \in Z_n\}$ to be the basis vectors of an orthonormal wavelet transform of length n with J decomposition levels. We assume periodic extension is used to handle the boundary. Suppose that $\boldsymbol{v}_0$ corresponds to the single scaling coefficient or the mean value, then all other vectors must have zero mean (see Lemma 1). Denote $c^{\{k\}}$ be the vector with k entries, all equal to c. Consider the following p vectors defined in $\mathbb{R}^p$

$$
\begin{aligned}
\boldsymbol{w}_0 &= (1^{\{p\}}) \,/\, s_0 \\
\boldsymbol{w}_1 &= (1^{\{p-1\}}, -p+1) \,/\, s_1 \\
\boldsymbol{w}_2 &= (1^{\{p-2\}}, -p+2, 0) \,/\, s_2 \\
&\cdots \\
\boldsymbol{w}_{p-n} &= (1^{\{p-n\}}, -p+n, 0^{\{n-1\}}) \,/\, s_{p-n} \\
\boldsymbol{w}_{p-n+1} &= (\boldsymbol{v}_1, 0^{\{p-n\}}) \\
&\cdots \\
\boldsymbol{w}_{p-1} &= (\boldsymbol{v}_{n-1}, 0^{\{p-n\}}).
\end{aligned}
$$

Here s_k is the scale factor such that $\|\boldsymbol{w}_k\| = 1$. The orthogonality of the new set $\{\boldsymbol{w}_k : k \in Z_p\}$ can be easily verified given the fact that $\{\boldsymbol{v}_m : 1 \le m < n\}$ are orthonormal vectors with zero mean. Therefore, $\{\boldsymbol{w}_k : k \in Z_p\}$ is an orthonormal basis for $\mathbb{R}^p$ that satisfies the *Condition Z*. For a length p input vector $\boldsymbol{x} = (x_0, x_1, \ldots, x_{p-1})$, the transform coefficients correspond to $\boldsymbol{w}_k$, where $p-n \le k \le p-1$, can be computed efficiently using the usual DWT with J levels on the first n samples $\boldsymbol{x}' = (x_0, x_1, \ldots, x_{n-1})$. The last scaling coefficient is then replaced by $p-n+1$ coefficients corresponding to the basis vectors $\boldsymbol{w}_k,\ k = 0, \ldots, p-n$. Thus the new basis in $\mathbb{R}^p$ also has fast transforms.

References

[1] M. Vetterli and J. Kovačević, *Wavelets and Subband Coding*, Prentice-Hall, Englewood Cliffs, NJ, 1995.

[2] D. L. Donoho, M. Vetterli, R. A. DeVore, and I. Daubechies, "Data compression and harmonic analysis," *IEEE Trans. Inform. Th.*, vol. 44, no. 6, pp. 2435–2476, October 1998.

[3] S. Mallat, *A Wavelet Tour of Signal Processing*, Academic Press, 2nd edition, 1999.

[4] K. R. Rao and P. Yip, *Discrete Cosine Transform: Algorithms, Advantages, Applications*, Academic Press, 1990.

[5] E. J. Candès, *Ridgelets: Theory and Applications*, Ph.D. thesis, Department of Statistics, Stanford University, 1998.

[6] E. J. Candès and D. L. Donoho, "Ridgelets: a key to higher-dimensional intermittency?," *Phil. Trans. R. Soc. Lond. A.*, pp. 2495–2509, 1999.

[7] S. R. Deans, *The Radon Transform and Some of Its Applications*, John Wiley & Sons, 1983.

[8] D. L. Donoho and M. R. Duncan, "Digital curvelet transform: strategy, implementation and experiments," in *Proc. Aerosense 2000, Wavelet Applications VII.* SPIE, 2000, vol. 4056, pp. 12–29.

[9] J. L. Starck, E. J. Candès, and D.L. Donoho, "The curvelet transform for image denoising," *IEEE Trans. Image Proc.*, vol. 11, pp. 670–684, Jun. 2002.

[10] A. Averbuch, R. R. Coifman, D. L. Donoho, M. Israeli, and J. Waldén, "Fast slant stack: A notion of Radon transform for data in a Cartesian grid which is rapidly computable, algebraically exact, geometrically faithful and invertible," Tech. Rep., 2001, Preprint, `http://www.math.tau.ac.il/~amir1/`.

[11] E. D. Bolker, "The finite Radon transform," in *Integral Geometry (Contemporary Mathematics, Vol. 63)*, S. Helgason R. L. Bryant, V. Guillemin and R. O. Wells Jr., Eds., pp. 27–50. 1987.

[12] S. Mallat and W. L. Hwang, "Singularity detection and processing with wavelets," *IEEE Transactions on Information Theory, Special Issue on Wavelet Transforms and Multiresolution Signal Analysis*, vol. 38, no. 2, pp. 617–643, March 1992.

[13] G. T. Herman, *Image Reconstruction from Projections: The Fundamentals of Computerized Tomography*, Academic Press, 1980.

[14] A. Rosenfeld and A. C. Kak, *Digital Picture Processing*, Academic Press, 2nd edition, 1982.

[15] R. M. Mersereau and A. V. Oppenheim, "Digital reconstruction of multidimensional signals from projections," *Proc. of IEEE*, vol. 62, no. 10, pp. 1319–1338, Oct 1974.

[16] G. Beylkin, "Discrete Radon transform," *IEEE Trans. Acoust. Speech Sig. Proc.*, vol. 35, pp. 162–172, 1987.

[17] J. L. C. Sanz, E. B. Hinkle, and A. K. Jain, *Radon and Projection Transform-Based Computer Vision*, Springer Verlag, 1988.

[18] B. T. Kelley and V. K. Madisetti, "The fast discrete Radon transform – I: Theory," *IEEE Trans. on Image Proc.*, vol. 2, pp. 382–400, Jul 1993.

[19] I. Gertner, "A new efficient algorithm to compute the two-dimensional discrete Fourier transform," *IEEE Trans. Acoust. Speech Sig. Proc.*, vol. 36, pp. 1036–1050, Jul 1988.

[20] F. Matúš and J. Flusser, "Image representation via a finite Radon transform," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 15, no. 10, pp. 996–1006, Oct 1993.

[21] P. M. Salzberg and R. Figueroa, "Tomography on the 3D-torus and crystals," in *Discrete Tomography: Foundations, Algorithms and Applications*, G. T. Herman and A. Kuba, Eds., pp. 417–434. Birkhäuser, 1999.

[22] N. D. Vvedenskaya and S. G. Gindikin, "Discrete Radon transform and image reconstruction," in *Mathematical problems of tomography (Translations of mathematical monographs, Vol. 81)*, I. M. Gelfand and S. G. Gindikin, Eds., pp. 141–188. American Mathematical Society, 1990.

[23] R. Lidl and H. Niederreiter, *Introduction to Finite Fields and Their Applications*, Cambridge University Press, revised edition, 1994.

[24] I. Daubechies, *Ten Lectures on Wavelets*, SIAM, Philadelphia, PA, 1992.

[25] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge University Press, 1985.

[26] V. K. Goyal, J. Kovačević, and J. A. Kelner, "Quantized frame expansions with erasures," *Applied and Computational Harmonic Analysis*, vol. 10, no. 3, pp. 203–233, May 2001.
[27] C. M. Brislawn, "Classification of nonexpansive symmetric extension transforms for multirate filter banks," *Applied and Computational Harmonic Analysis*, vol. 3, pp. 337–357, 1996.
[28] C. Herley and M. Vetterli, "Wavelets and recursive filter banks," *IEEE Trans. Signal Proc.*, vol. 41, no. 8, pp. 2536–2556, August 1993.
[29] D. Donoho and I. Johnstone, "Ideal spatial adaptation via wavelet shrinkage," *Biometrika*, pp. 425–455, December 1994.
[30] E. J. Candès and D. L. Donoho, "Curvelets – a suprisingly effective nonadaptive representation for objects with edges," in *Curve and Surface Fitting*, A. Cohen, C. Rabut, and L. L. Schumaker, Eds., Saint-Malo, 1999, Vanderbilt University Press.

Minh N. Do was born in Thanh Hoa, Vietnam, in 1974. He received the B.Eng. degree in computer engineering (with first class honors) from the University of Canberra, Australia, in 1997, and the Dr.Sci. degree in communication systems from the Swiss Federal Institute of Technology Lausanne (EPFL), Switzerland, in 2001.
Since 2002, he has been an Assistant Professor with the Department of Electrical and Computer Engineering and a Research Assistant Professor with the Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign. His research interests include wavelets and harmonic analysis, multirate systems, image and multidimensional signal processing, and visual information representation.
Dr. Do received a Silver Medal from the 32nd International Mathematical Olympiad in Sweden (1991) and a University Medal from the University of Canberra (1997).

Martin Vetterli received the Dipl. El.-Ing. degree from ETH Zürich (ETHZ), Switzerland, in 1981, the MS degree from Stanford University in 1982, and the Doctorat ès Science degree from EPF Lausanne (EPFL), Switzerland, in 1986.
He was a Research Assistant at Stanford and EPFL, and has worked for Siemens and AT&T Bell Laboratories. In 1986, he joined Columbia University in New York where he was last an Associate Professor of Electrical Engineering and co-director of the Image and Advanced Television Laboratory. In 1993, he joined the University of California at Berkeley where he was a Professor in the Dept. of Electrical Engineering and Computer Sciences until 1997, and holds now an Adjunct Professor position. Since 1995, he is a Professor of Communication Systems at EPF Lausanne, Switzerland, where he chaired the Communications Systems Division (1996/97), and heads of the Audio-Visual Communications Laboratory. He held visiting positions at ETHZ (1990) and Stanford (1998).
He is a fellow of the IEEE, a member of SIAM, and was the Area Editor for Speech, Image, Video, and Signal Processing of the IEEE Transactions on Communications. He is also on the editorial boards of Annals of Telecommunications, Applied and Computational Harmonic Analysis and The Journal of Fourier Analysis and Applications. He received the Best Paper Award of EURASIP in 1984 for his paper on multidimensional subband coding, the Research Prize of the Brown Bovery Corporation (Switzerland) in 1986 for his doctoral thesis, the IEEE Signal Processing Society's Senior Award in 1991 and in 1996 (for papers with D.LeGall and K.Ramchandran, respectively). He received the Swiss National Latsis Prize in 1996, the SPIE Presidential award in 1999, and the IEEE Signal Processing Technical Achievement Award in 2001. He is a member of the Swiss Council on Science and Technology. Since 2001, he directs the National Competence Center in Research on mobile information and communication systems (www.terminodes.org).
He was a plenary speaker at various conferences (e.g. 1992 IEEE ICASSP) and is the co-author, with J.Kovacevic, of the book **Wavelets and Subband Coding** (Prentice-Hall, 1995). He has published about 85 journal papers on a variety of topics in signal and image processing and holds 7 patents.
His research interests include wavelets, multirate signal processing, computational complexity, signal processing for telecommunications, digital video processing and compression and wireless video communications.