Discrete Ridgelet Transforms
for Image Representation

Minh N. Do[*] and Martin Vetterli[†]

Audio-Visual Communications Laboratory (LCAV)

Swiss Federal Institute of Technology (EPFL)

1015 Lausanne, Switzerland

Email: {Minh.Do, Martin.Vetterli}@epfl.ch; Web: lcavwww.epfl.ch

Abstract

The ridgelet transform (Candès and Donoho, 1999) was introduced as a new multiscale representation for functions on continuous spaces that are smooth away from discontinuities along lines. In this paper, we present several discrete versions of the ridgelet transform that lead to algorithmic implementations. The resulting transforms are invertible, non-redundant and computed via fast algorithms. Furthermore, this construction leads to a family of directional orthonormal bases for digital images. Numerical results show that the new transforms are 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 denoising.

[*] Corresponding author. Address: see above; Phone: +41 21 693 7663; Fax: +41 21 693 4312.
[†] Also with Department of EECS, UC Berkeley, Berkeley CA 94720, USA.

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] and the discrete cosine transform [3].

The success of wavelets are mainly due to their 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. Intuitively, wavelets in two dimensions obtained by a tensor-product of one dimensional wavelets will be good at isolating the discontinuity across the edge, but will not see the smoothness along the edge.

This fact has a direct impact on the performance of many applications that use wavelets. As an example, for the image denoising problem, state-of-the-art techniques are based on thresholding of wavelet coefficients [4], [5]. Despite their simplicity, these methods work very effectively, mainly due to the property of the wavelet transform that most image information is contained in a small number of significant coefficients – around the locations of singularities or image edges. However, since wavelets fail to represent efficiently singularities along lines or curves, wavelet-based techniques fail to explore the geometrical structure that is typical smooth edges in images. Therefore, new denoising schemes which are based on "true" two-dimensional (2-D) transforms are expected to improve the performance over the current wavelet-based methods.

To overcome the weakness of wavelets in higher dimensions, Candès and Donoho [6], [7] recently pioneered a new system of representations named *ridgelets* which they showed to deal effectively with *line* singularities in 2-D. The idea is to map a line singularity into a point singularity using the Radon transform [8]. Then, the wavelet transform can be used to handle the point singularity. 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 the polar coordinate or in the rotation operator, and thus the resulting transforms would be either redundant or can not be perfectly reconstructed. The approaches described in [9], [10], which are based on interpolations over the "concentric-squares" grid in the Fourier plane, lead to discrete ridgelet transforms with exact reconstruction, but they have a

factor two overcompleteness.

In this paper, we propose several versions of the discrete ridgelet transform that achieve both the invertible and non-redundancy requirements. A key step in our method is an invertible discrete Radon transform. This construction leads to a new family of orthonormal bases for digital images. Its properties are demonstrated and studied in several applications.

Back to the image denoising problem, there are other approaches that explores the geometrical regularity of edges, for example by chaining adjacent wavelet coefficients and then thresholding them over those contours [11]. However, the discrete ridgelet transform approach, with its "built-in" linear geometrical structure, provide a more elegant 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 finite Radon transform is also studied in the frame theory. 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 some variations on the initial design of the finite ridgelet transform in order to reduce artifacts. In Section VI, we extend the finite ridgelet idea further to obtain a transform that is free from periodization but at the cost of being redundant. Numerical experiments are presented in Section VII, where the new transforms are compared with the traditional ones, especially wavelets. We conclude in Section VIII 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 [6], [7]

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

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

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

Fig. 1 plots a typical ridgelet: the function is oriented at the angle θ and is constant along the lines $x_1 \cos \theta + x_2 \sin \theta = const$.

In comparison, the (separable) continuous wavelet transform (CWT) in $\mathbb{R}^2$ of $f(\boldsymbol{x})$ can be
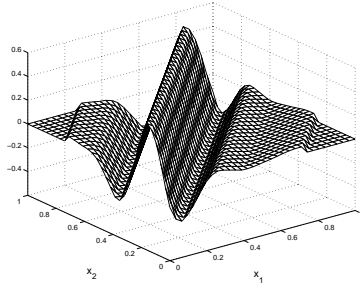
3

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

written as

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

where the wavelets in 2-D are tensor products

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

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

The CRT appears 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 at representing objects with isolated point singularities, while ridgelets are very effective at representing objects with singularities along lines. In fact, one can loosely view ridgelets as a way of concatenating 1-D wavelets along lines. Hence the motivation for using ridgelets in image processing tasks is very appealing as 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

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

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

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

---

[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.
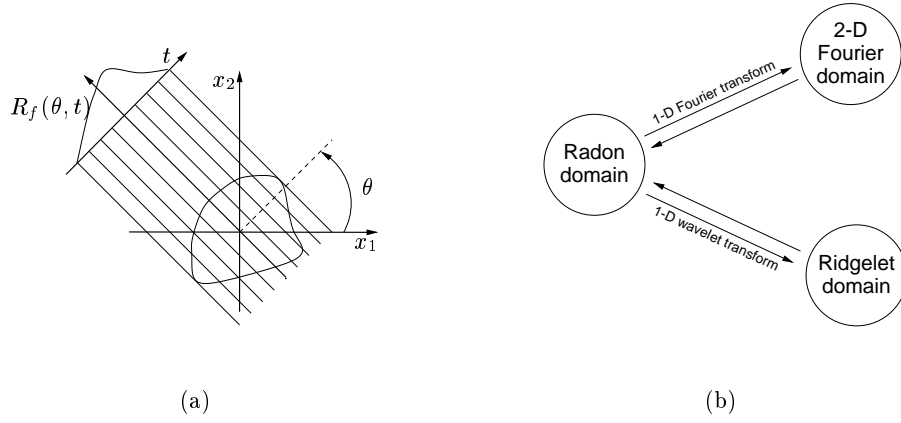
Fig. 2.   (a) The Radon transform in $\mathbb{R}^2$. For a fixed θ, $R_f(\theta, t)$ is a slice or projection in the Radon domain. (b) Relations between transforms. The ridgelet transform is the application of 1-D wavelet transform to the slices of the Radon transform, while the 2-D Fourier transform is the application of 1-D Fourier transform to those Radon slices.

Fig. 2(a) shows a graphical representation of the Radon transform. It is instructive to note that if in (6) instead of taking a 1-D wavelet transform, the application of 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

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

This is the famous *projection-slice* theorem and is used often in image reconstruction from projection methods [12], [13]. The relations between the various transforms are shown in Fig. 2(b).

III. Finite Radon Transform

A. Forward and Inverse Transforms

As suggested in the previous section, a discrete ridgelet transform can be obtained via a discrete Radon transform. Numerous discretizations of the Radon transforms have been devised to approximate the continuous formula [12], [13], [14], [15], [16]. However to our knowledge, none of them were specially designed to be invertible transforms for digital images. Alternatively, the *finite*[2] Radon transform theory [17], [18], [19], [20], which originated from combinatorics, provides an interesting solution. Also, in [21], 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

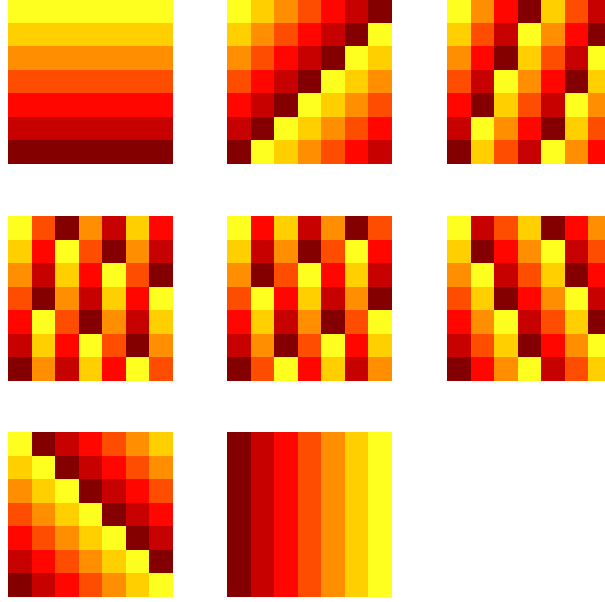[2]That means a transform for *finite* length signals.

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.

continuous Radon transform in the Euclidean geometry. Denote $Z_p = \{0, 1, \ldots, p-1\}$, where p is a prime number. Note that Z_p is a finite field with modulo p operations [22]. For later convenience, we denote $Z_p^* = \{0, 1, \ldots, p\}$.

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

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

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

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

Fig. 3 displays an example of the finite lines $L_{k,l}$ where points in the grid Z_p^2 are represented by image pixels.

We observe that in the FRAT domain, the energy is best compacted if the mean is subtracted from the image $f[i, j]$ previous 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 are in 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

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

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).

In 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

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

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

$$P_{i,j} = \{(k, l) : l = j - ki \quad (\mathrm{mod}\ p), k \in Z_p\} \cup \{(p, i)\} \tag{12}$$

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

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

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.

Both the forward and inverse of the FRAT require p^3 additions and p^2 multiplications. In [19], Matúš and Flusser described a fast implementation of the FRAT in which each projection needs to pass through every pixel of the original image once using p histogrammers, one for each FRAT coefficient of that projection. With our implementation by programming in C on a Sun Ultra 5 computer, both the forward and inverse FRAT take less than 0.1 second to compute for an image of size 127×127.

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 the 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

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

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

This is in analogy with the line equation: $x_1 \cos \theta + x_2 \sin \theta - t = 0$ in $\mathbb{R}^2$. So for a finite line defined as in (14), (a, b) has the role as 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.

To relate (9) with (14), consider the general line equation in (14)

$$ai + bj - t = 0. \tag{15}$$

If $b \neq 0$ then, (15) $\Leftrightarrow j = -b^{-1}ai + b^{-1}t$, where b^{-1} denotes the multiplicative inverse of b in the finite field Z_p, i.e. $bb^{-1} = 1 \ (\text{mod } p)$. Otherwise, if $b = 0$, then it is necessary that $a \neq 0$, thus (15) $\Leftrightarrow i = a^{-1}t$.

So by matching the line equations in (9) and (14), we have the following equivalences between these two specifications of finite lines in Z_p^2:

$$\begin{aligned}
L'_{a,b,t} &\equiv L_{k,l} \quad \text{if } a = -kb, b \neq 0, t = bl, \quad \text{for } 0 \leq k < p, \quad \text{and} \\
L'_{a,b,t} &\equiv L_{p,l} \quad \text{if } a \neq 0, b = 0, t = al.
\end{aligned}$$

In other words, there is a many (exactly $p-1$) to one mapping between the line parameters in (14) and in (9), such that they represent the same line. It is easy to see that for any $c \in Z_p, c \neq 0$ then $\{cl : l \in Z_p\}$ is the same as the set Z_p. So for a fixed normal vector (a, b), the set of parallel lines $\left\{ L'_{a,b,t} : t \in Z_p \right\}$ equals to the set of p lines $\{L_{k,l} : l \in Z_p\}$ with a 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 also 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

$$r_{a,b}[t] = FRAT_f[a, b, t] = \frac{1}{\sqrt{p}} \sum_{(i,j) \in L'_{a,b,t}} f[i,j]. \tag{16}$$

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 an 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.

8

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

$$
\begin{aligned}
\boldsymbol{u}_k &= (-k, 1) \quad \text{for } k = 0, 1, \ldots, p-1, \text{ and} \\
\boldsymbol{u}_p &= (1, 0).
\end{aligned} \tag{17}
$$

In order to provide a complete representation, we need the FRAT to be defined as in (16) 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 which is the $p+1$ optimal normal vectors for 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 [19].

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

$$
F[u, v] = \frac{1}{p} \sum_{(i,j) \in Z_p^2} f[i, j] W_p^{ui+vj}, \tag{18}
$$

and for FRAT projections on Z_p as

$$
R_{a,b}[w] = \frac{1}{\sqrt{p}} \sum_{t \in Z_p} r_{a,b}[t] W_p^{wt}. \tag{19}
$$

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):

$$
R_{a,b}[w] = F[aw, bw]. \tag{20}
$$

Proof: Substituting (16) into (19) 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

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

$\square$

From (20), 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
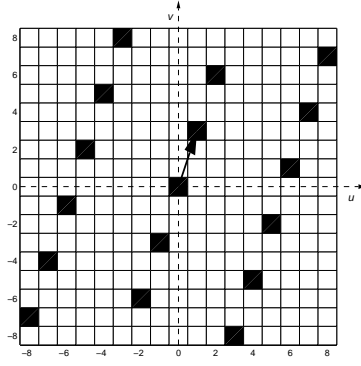
Fig. 4. Example of a discrete Fourier slice (consisted of all the black squares) with the best normal vector for that FRAT projection. In this example, $p = 17$ and the slope $k = 11$. The normal vector can be chosen as a vector from the origin to any other points on the Fourier slide. The best normal vector is $(1, 3)$ (the solid arrow).

(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.

Fig. 4 illustrates this by showing an example of discrete Fourier slice. The normal vector for the corresponding FRAT projection can be chosen as a vector from the origin to *any* other points on the Fourier slice. However, the best normal vector is selected as the closest point to the origin.

Furthermore, 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 least "wrap around", which is due to the periodization in the transform. The effect of the new ordering of FRAT coefficient in the image domain is illustrated in Fig. 5 for the same example projection. As can be seen, the "wrap around" effect is significantly reduced with the optimal ordering compared with 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

$$(a_k^\star, b_k^\star) = \arg \min_{\substack{(a_k, b_k) \in \{ n\mathbf{u}_k \, : \, 1 \leq n \leq p-1 \} \\ \text{s.t.} \ \ \mathcal{C}_p(b_k) \geq 0}} \| (\mathcal{C}_p(a_k), \mathcal{C}_p(b_k)) \| . \tag{21}$$

Here $\mathcal{C}_p(x)$ denotes the centralized function of period p: $\mathcal{C}_p(x) = x - p.\text{round}(x/p)$. 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 (21). Fig. 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 (17), the new set $\left\{ (a_k^\star, b_k^\star) : k \in Z_p^* \right\}$ provides a much more uniform angular coverage.

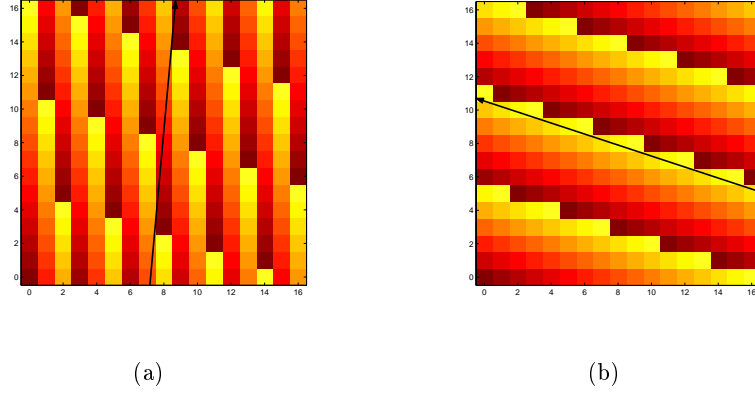(a)   (b)

Fig. 5. Lines for the FRAT projection as shown in Fig. 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.

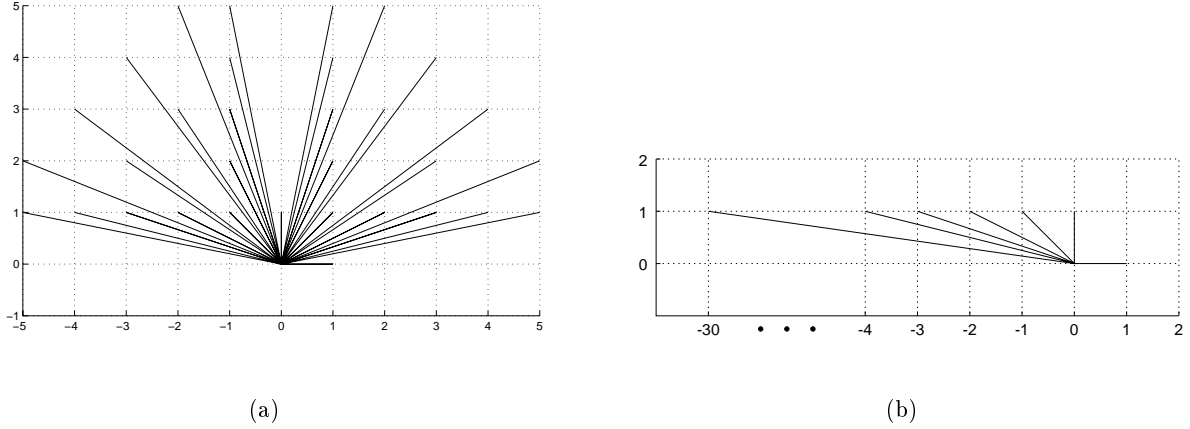
(a)   (b)

Fig. 6. The set of normal vectors, which indicate the represented directions, for the FRAT of size $p = 31$ using: (a) optimal ordering; (b) usual ordering.

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 as described before. For a given p, solving (21) requires $O(p^2)$ operations and therefore it is negligible compared with the transforms themselves. Furthermore, this can be pre-computed, thus only presents as a "one-time" cost.

For the sake of simplicity and with an abuse of notation, in the sequel we write $r_k[t]$ for $r_{a_k^\star, b_k^\star}[t]$. 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 $0 \leq k \leq p,\ 0 \leq t < p$.

C. Frame Analysis of the FRAT

Since the FRAT is an invertible linear operator, it can be considered as a *frame* – a useful concept in analyzing redundant expansions [23]. In this section we will study the FRAT in more detail and reveal some of its interesting properties in the frame setting.

11

For a detailed introduction to frames, readers are referred to [5] (Chapter 5). 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. It can be shown that any finite set of vectors that spans $\mathbb{R}^N$ is a frame. Associated with a frame is the *frame operator* $\mathcal{F}$, defined as the linear operator from $\mathbb{R}^N$ to $\mathbb{R}^M$

$$(\mathcal{F}x)_n = \langle x, \varphi_n \rangle, \qquad \text{for } n = 1, \ldots, M. \tag{23}$$

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

$$x^T A x \leq x^T F^T F x \leq x^T B x, \qquad \forall x \in \mathbb{R}^N. \tag{24}$$

Since $F^T F$ is symmetric, it is diagonalizable in an orthonormal basis [24], thus (24) 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 $F^T F = A \cdot I_N$, where I_N denotes the $N \times N$ identity matrix; which means the transpose of F equals its left inverse within a scale factor.

Now let us return to the FRAT; it can be regarded as a frame operator in $l_2(Z_p^2)$ with the frame $\{\varphi_{k,l}\}$ defined as

$$\varphi_{k,l} = p^{-1/2} \delta_{L_{k,l}} \quad k \in Z_p^*, l \in Z_p, \tag{25}$$

where δ_S denotes the characteristic function for a set S in Z_p^2.

Note that this frame is normalized since $\left\| \varphi_{(k,l)} \right\| = 1$. Moreover, using the properties of lines in the finite geometry Z_p^2, it is easy to verify that

$$\langle \varphi_{k,l}, \varphi_{k',l'} \rangle = \begin{cases} 1 & \text{if } k = k', l = l' \\ 0 & \text{if } k = k', l \neq l' \\ 1/p & \text{if } k \neq k' \end{cases} \tag{26}$$

Hence, the minimum angle between any two frame vectors of the FRAT is: $\cos^{-1}(1/p)$, which approaches the right angle for large p. So we can say that the FRAT frame is "almost" orthogonal.

Writing images as column vectors, the FRAT can be regarded as a left matrix multiplication with $p^{-1/2}R$, 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.

Similarly, 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.[3] Since the FBP is an inverse of the FRAT for the subspace of zero-mean images defined on $\mathbb{Z}_p^2$, the FRAT is a *normalized tight frame* in this subspace. Thus we obtain the following result.

Proposition 1: If f is a zero-mean image on Z_p^2, then

$$f = \sum_{k=0}^{p}\sum_{l=0}^{p-1}\langle f, \varphi_{k,l}\rangle \varphi_{k,l}. \tag{27}$$

For general images, the FRAT operator is not a tight frame as shown in the following proposition.

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

Proof: From discussion before, these tightest bounds can be computed from the eigenvalues of $C = p^{-1}R^T R$. Since R is the incidence matrix for lines in Z_p^2, $(R^T R)_{(i,j),\ (i',j')}$ equals to the number of lines that goes 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 $B = p + 1$. □

---

[3]In a more technical term, FBP is the adjoint of FRAT.

Remark: It is instructive to note that constant images on Z_p^2 are eigenvectors of C with 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.

What is the consequence of those frame bounds? The answer is in the robustness of the reconstruction when there is noise due to nonlinear approximation and quantization. In [23], it is shown that with a simple additive noise model for quantization, a tight frame is optimal among normalized frames in minimizing mean-squared error. Therefore by subtracting the mean out of the image before applying the FRAT (which leads to a tight frame) we make the FRAT coefficients more robust against processing noise.

IV. Orthonormal Finite Ridgelet Transform

Now with an invertible FRAT, 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. The overall result can be called *finite ridgelet transform* (FRIT). Fig. 7 depicts these steps.
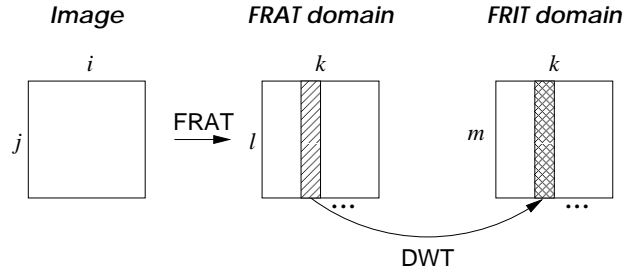

Fig. 7. Diagram for the FRIT. After taking the FRAT, a DWT is applied to each of the FRAT slices or projections where k is fixed.

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

$$G_0^{(J)}(z) = \prod_{k=0}^{J-1} G_0(z^{2^k}),$$

$$G_1^{(j)}(z) = G_1(z^{2^{j-1}}) \prod_{k=0}^{j-2} G_0(z^{2^k}), \quad j = 1, \dots, J.$$

The basis functions from $G_0^{(J)}$ are called the scaling functions where all the others functions in the wavelet basis are called wavelet functions. Normally, 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, $G_1^{(j)}(z)|_{z=1} = 0$, $\forall j = 1, \dots, J$, which means all wavelet basis functions have zero sum.

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

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

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, $\boldsymbol{w}_m$, $m = 1, \dots, p-1$, have zero sum.

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

Conversely, let us assume that each basis function $\boldsymbol{w}_m$, $1 \leq m \leq p-1$, has zero sum. 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 $S_0 = \{c\boldsymbol{1} : c \in \mathbb{R}\}$. We have $\langle c\boldsymbol{1}, \boldsymbol{w}_m \rangle = c \sum_l w_m[l] = 0$, $\forall m = 1, \dots, p-1$, 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. $\qquad\square$

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).

15

By definition, the FRIT can now be written as

$$
\begin{aligned}
FRIT_f[k,m] &= \langle FRAT_f[k,\bullet], w_m^{(k)}[\bullet] \rangle \\
&= \sum_{l \in Z_p} w_m^{(k)}[l] \langle f, \varphi_{k,l} \rangle \\
&= \langle f, \sum_{l \in Z_p} w_m^{(k)}[l] \ \varphi_{k,l} \rangle. \tag{28}
\end{aligned}
$$

Here $\{\varphi_{k,l}\}$ is the FRAT frame which is defined in (25). Hence we can write the basis functions for the FRIT as follows:

$$
\rho_{k,m} = \sum_{l \in Z_p} w_m^{(k)}[l] \ \varphi_{k,l}. \tag{29}
$$

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 (26), 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, $k \neq k'$, again using (26) we obtain

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

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 direction 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 (at any direction) and denote it as ρ_0. Hence we have proved the following theorem.

Theorem 2: Given $p+1$ orthonormal bases in $l^2(Z_p)$ (which can be the same): $\left\{ \boldsymbol{w}_m^{(k)} : m \in Z_p \right\}$, $0 \leq k \leq p$, 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 (29).

Remarks

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 lead 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 Fourier-type transform 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 would 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.

Fig. 8 displays some of basis images for the 7×7 FRIT using 1-D Haar wavelets. As can be seen from the figure, FRIT basis images have elongated linear structure which closely resemble the continuous ridge functions.
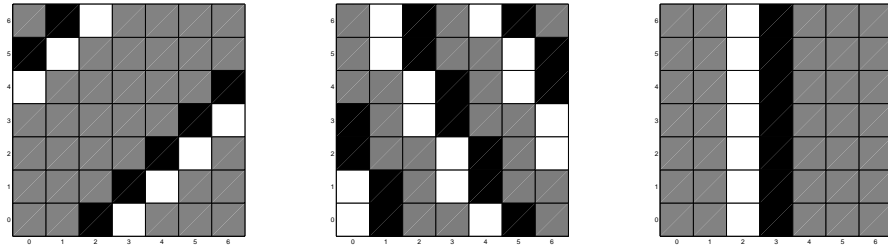

Fig. 8. Examples of 7×7 FRIT basis functions using Haar wavelets. Black, gray and white pixels correspond respectively to negative, zero and positive values. Notice the "wrap-around" effect due to the periodic property of the FRAT.

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 employ a similar strategy as in the block cosine transform by extending the image symmetrically about its borders.
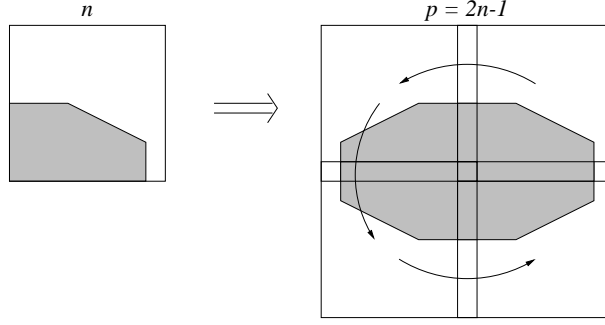

Fig. 9. Extending the image symmetrically about its borders in order to reduce the discontinuities across the image borders due to the periodization.

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 $f[i, j]$, $0 \leq i, j < n$. 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 Fig. 9)

$$\check{f}[i, j] = f[|i|, |j|], \quad -n < i, j < n. \tag{30}$$

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 indexes 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 (30), 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

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

18

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

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

$$\check{R}_{a,b}[w] = \check{R}_{|a|,|b|}[w]. \tag{32}$$

From (31) 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 (32) reveals the duplications among those projections. In fact, with the set of optimal normal vectors in (21), 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 Fig. 6). Furthermore, as in (10), the redundancy among the projections of the folded FRAT can be written as

$$\check{r}_{a_k^\star, b_k^\star}[0] + 2 \sum_{t=1}^{n-1} \check{r}_{a_k^\star, b_k^\star}[t] = \frac{1}{\sqrt{p}} \sum_{-n < i,j < n} \check{f}[i, j]. \tag{33}$$

The next proposition summaries the above results.

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

$$\check{r}_{a_k^\star, b_k^\star}[t] \qquad \text{such that } \mathcal{C}_p(a_k^\star) \geq 0 \text{ and } 0 < t < n, \tag{34}$$

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 (34) 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 overlaps. 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 [25] or orthogonal symmetric IIR wavelets [26].

B. Multilevel FRIT's

In the FRIT schemes 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 describe a different scheme where the number of directions can be controlled, plus the basis functions have smaller support. Assume that the input image has 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. 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.

C. Generalized (or Galois) Finite Radon and Ridgelet Transforms

All algebraic properties of the FRAT defined in Section III are based on the fact that Z_p is a finite field with the modulo p addition and multiplication. Therefore one could extend this and consider a general (or Galois) finite Radon transform (GFRAT) that uses more general finite fields $GF(p^r)$ [22].

Denote $n = p^r$ where p is a prime and r is a positive integer then we can define the set of "lines" on the affine plane Z_n^2 exactly in the same way in (9) or (14) except that modulo p operations are replaced by the finite field $GF(p^r)$ operations.

For the GFRAT, the issue of ordering coefficients is even more complex since there is no "natural" ordering for elements of a general finite field. Besides, a disappointing aspect of the GFRAT is its lines has little resemble to the natural lines in images except for the horizontal and vertical lines. Thus there is little hope of using GFRAT as efficient representation for images with linear discontinuities. Nevertheless, by applying any orthogonal DWT on the projections of GFRAT in the way described in Section IV, we would obtain an orthogonal transform for digital images. In addition, wavelet transforms defined over finite fields [27], [28] could be used. Potential applications for the resulting transforms could be in constructing of communication codes or cryptography.

VI. Ridgelet Frames

As mentioned above, the orthogonal FRIT has a drawback which is due to the annoying "wrap-around" effect of the FRAT. As a result, in non-linear approximation, large FRIT coefficients that respond to a straight edge may also produce "alias" edges in the reconstructed image. Furthermore, the construction of FRAT creates "holes" along some of its represented lines (refer to Fig. 3). In this section, we will describe a construction for the discrete Radon transform that overcomes these weaknesses and imitates more closely the continuous transform. However, as we will see the price to pay is redundancy, and thus we get ridgelet frames.

The central issue is to form a collections of lines, which themselves are subsets of pixels, so that the Radon transform can be defined as summations of pixels on each of those lines. To remove the periodic effect of the FRAT, we need to drop the modulo operations in the finite line equations $ai + bj = t$. In particular, we want to define lines on a finite grid Z_n^2 as

$$\left\{ (i,j) \in Z_n^2 : q_t \leq ai + bj < q_{t+1} \right\}. \tag{35}$$

So it remains to construct a set of normal vectors (a, b) and translation steps q_t. The set of normal vectors will control the angular sampling or represented directions for the discrete Radon projections, while the translation steps will control the radical sampling on each projection.

For the normal vectors, given a positive integer m, we construct $4m$ normal vectors as follows (also see Fig. 10):

$$(a_k, b_k) = \begin{cases} (m, k) & 0 \leq k < m \\ (2m - k, m) & m \leq k < 3m \\ (-m, 4m - k) & 3m \leq k < 4m \end{cases} \tag{36}$$
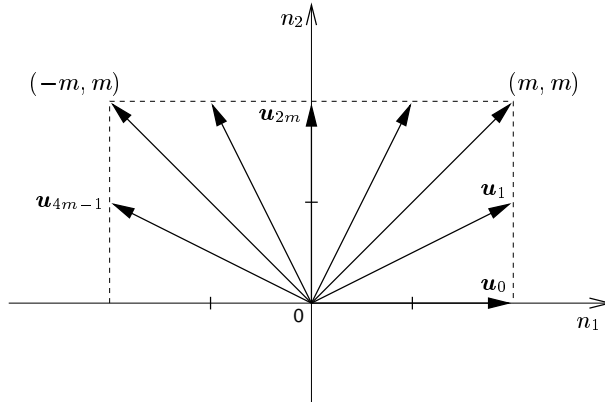

Fig. 10. Set of directions for the discrete Radon transform.

The directions are defined in this way in order to provide an uniform "coverage" for discrete lines on a square image. Increasing m leads to finer angular sampling. As a result, the angles

21

for the Radon transform are quantized into a set of $4m$ values:

$$\theta_k = \tan^{-1}(b_k/a_k), \quad 0 = \theta_0 < \theta_1 < \ldots < \theta_{4m-1} < \pi.$$

For the translation steps, given a direction that is represented by a normal vector (a_k, b_k), based on the idea for representation of digital lines in [15], we define the "contour line image" $C_k[i, j]$ as follows:

$$C_k[i, j] = \text{round}\left(\frac{a_k i + b_k j}{|a_k| + |b_k|}\right) + d_k \tag{37}$$

where,

$$d_k = - \min_{i,j \in Z_n} \text{round}\left(\frac{a_k i + b_k j}{|a_k| + |b_k|}\right). \tag{38}$$

By construction, the range of $C_k[i, j]$ is precisely Z_n. Thus we can definite the subset of pixels that make up a discrete line with direction k and translation l as

$$S_{k,l} = \{(i, j) \ : \ C_k[i, j] = l\}. \tag{39}$$

These lines satisfy the condition in (35) and the resulting collection has the following properties:

1. There are exactly n lines for each directions. Thus the corresponding Radon transform can be arranged in a $n \times m$ matrix, one column for each projection.

2. For each direction, all n parallel lines have almost the same "width". They are non-overlapping and provide a complete coverage for Z_n^2.

3. Horizontal and vertical lines (when k equal to 0 and $2m$) are rows and columns of the images.

Fig. 11 shows an example collection of lines that is generated by this method. The lines imitate more closely the digital lines for image. With a set of lines $\{S_{k,l} : 0 \leq k < 4m, 0 \leq l < n\}$, the discrete Radon transform (DRAT) is defined as in the FRAT case:

$$DRAT_f[k, l] = \sum_{(i,j) \in S_{k,l}} f[i, j] = \langle f, \delta_{S_{k,l}} \rangle. \tag{40}$$

Since the lines for each projection are non-overlapped, a similar fast algorithm for the FRAT with n histogrammers can be used to compute the DRAT. The discrete ridgelet transform is obtained by applying the DWT on DRAT projections. The resulting ridgelet images clearly represent well linear singularities.

An obvious question is: Given the transform coefficients $DRAT_f[k, l]$, is it possible to reconstruct f in a numerically stable way? To answer this, we need to check the completeness and then the frame bounds of the set $\{\delta_{S_{k,l}} : 0 \leq k < 4m, \ 0 \leq l < n\}$ in $l_2(Z_n^2)$. Numerically, for a finite transform, the frame bounds can be computed as the squares of the smallest and largest
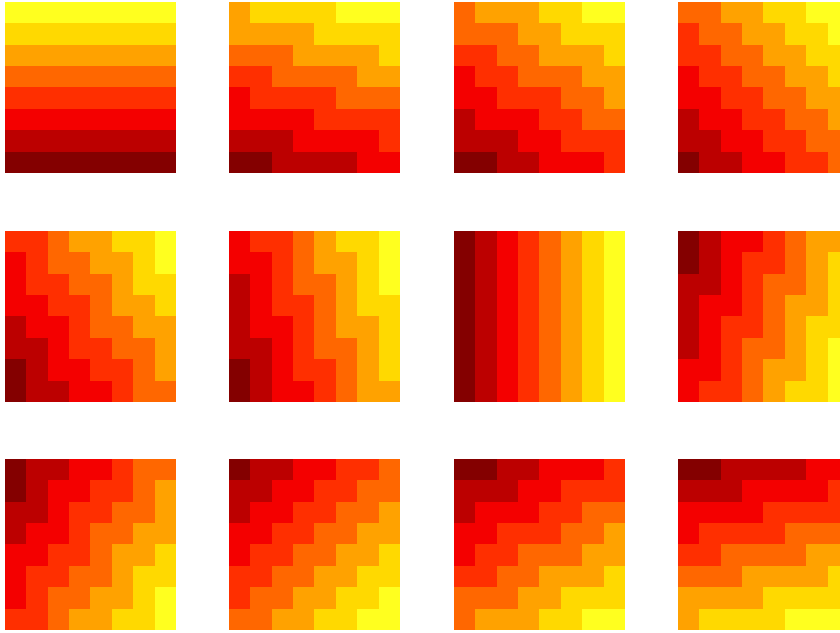
Fig. 11. Lines for the 8×8 discrete Radon frames. Compared with finite lines in Fig. 3, those resemble more the "natural" lines.

singular values of the transform matrix. Thus the frame bounds ratio is exactly the square of the condition number with respect to inversion of the transform.

Table I lists the computed results for some small size transforms. The constructed discrete Radon transform can provide complete representations for the image space with very small oversampling rates. However, the frame bounds ratios are large and they increase with the transform size. In general, there is a trade-off between oversampling and good conditioning for inversion.

Unfortunately, we can not find a fast structure algorithm for inversion of the discrete Radon frames. One possible solution is to resort to algebraic methods that use Kaczmarz's iterative algorithm or its variations for solving systems of linear equations [13], [29]. Alternatively, one could use this construction to produce a dictionary of ridgelet images and employ matching pursuits for image approximations [30].

VII. NUMERICAL EXPERIMENTS

A. Nonlinear Approximation

Following the study of the efficiency of the ridgelet transform in the continuous domain on the truncated Gaussian functions [7], we first perform numerical comparison on a 256×256 image of the function: $f(x_1, x_2) = 1_{\{x_2 < 2x_1 + 0.5\}} e^{-x_1^2 - x_2^2}$, using four 2-D transforms: DCT, DWT, FRAT

Image size (n)	Number of directions ($4m$)	Oversampling rate	Frame bounds ratio
4	8	1.56	7.24e+01
8	12	1.33	4.94e+03
	16	1.77	1.47e+03
16	20	1.18	2.72e+09
	24	1.41	1.17e+05
	28	1.64	1.35e+04
32	40	1.21	5.52e+06
	44	1.33	7.17e+05
	48	1.45	4.61e+05
	52	1.58	1.60e+05

TABLE I

CONDITIONS ON THE DISCRETE RADON FRAMES.

and FRIT.[4] Unless state otherwise, the discrete ridgelet transform used in the experiments of this section is the orthonormal FRIT. The wavelet used in both the DWT and FRIT is the "least asymmetric" orthogonal wavelet with 8-taps filters [31].

Our initial experiments indicate that in order to achieve good results, it is necessary to apply the DCT instead of the DWT to capture some strong periodic FRAT projections due to the "wrap around" effect (c.f. to the remarks at the end of Section IV). Without resorting to adaptive methods, we employ a simple, pre-defined scheme where the least "wrap around" FRAT projections – the ones with $\|(a_k^\star, b_k^\star)\| \leq D$ – use DWT, while all the others use DCT. We set $D = 3$ in our experiments, which means in the FRIT, only up to 16 FRAT projections are represented by the DWT.

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. Fig. 12 plots the results; clearly the FRIT achieves the best performance, as expected from the continuous theory.

Our next test is a real image of size 256×256 with straight edges. Fig. 13 displays the image together with the magnitude of its FRIT. In the FRIT image, each column corresponds to one direction. As can be seen, most of the FRIT coefficients are small.

To gain an insight into the FRIT, Fig. 14(a) plot the top five FRAT projections that contain

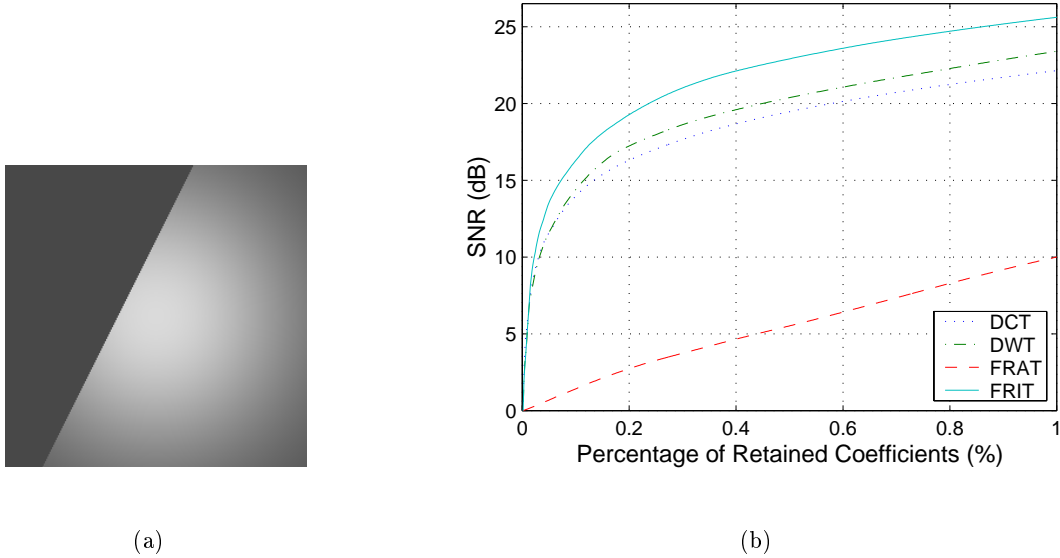[4]For the FRAT and FRIT, we need to extend the image size to the next prime number, 257.

(a)   (b)

Fig. 12.   (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, FRAT and FRIT.
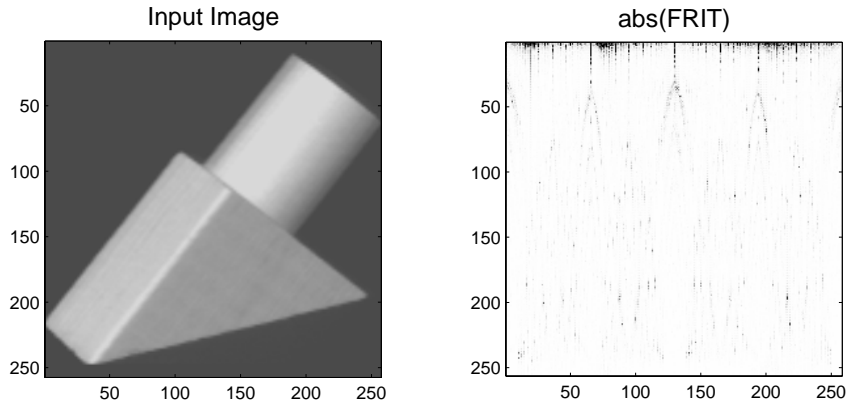

Fig. 13.   The FRIT coefficients of an image that is smooth away from straight edges are *sparse*. Large amplitude FRIT coefficients correspond to linear singularities in the image.

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. So we see that at first the FRAT compacts most of the energy of the image into a few projections (see Fig. 14(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 coefficients, whereas randomly located noisy
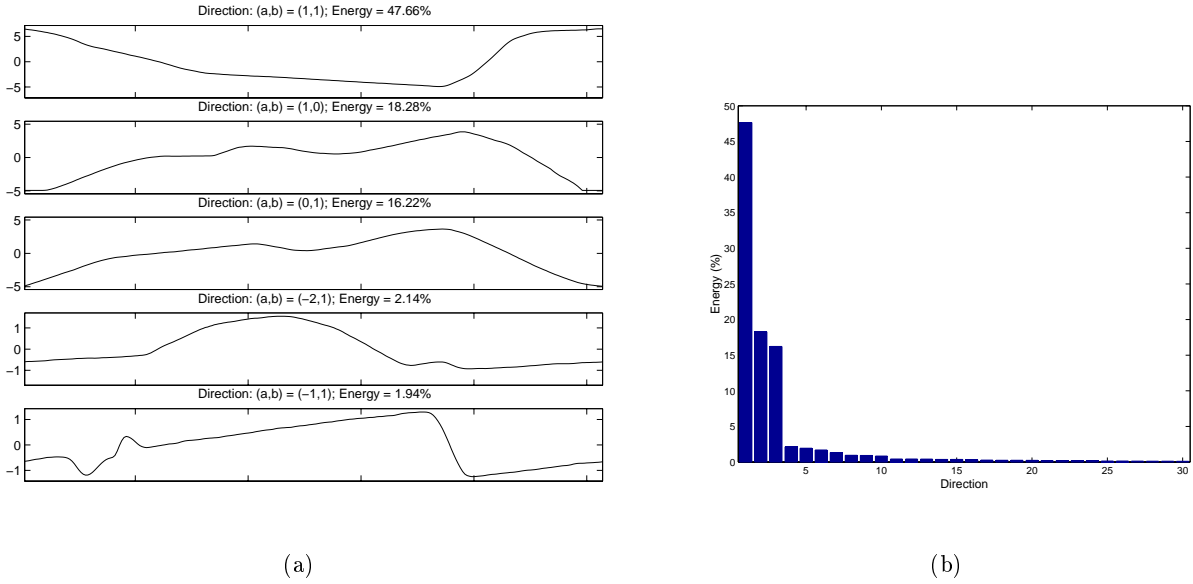
25

Fig. 14. (a) Top five FRAT projections of the image in Fig. 13 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.

singularities are unlikely to produce significant coefficients. By contrast, in the DWT domain, both image edges and noisy pixels can produce large 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 the 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 was proposed in [4] to the FRIT coefficients. More specifically, our denoising algorithm consists of the following steps:

1. Applying FRIT to the noisy image.

2. Hard-thresholding of FRIT coefficients with the universal threshold $T = \sigma\sqrt{2logN}$ where $N = p^2$ pixels.

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 their "wrap-around" effect 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. In some cases, this can enhances the visual appearance of the restored image.

4. (Optional) Adaptive Wiener filtering to reduce the "wrap-around" effect.

The above FRIT denoising algorithm is compared against the wavelet hard-thresholding method using the same threshold value. Fig. 15 displays 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.
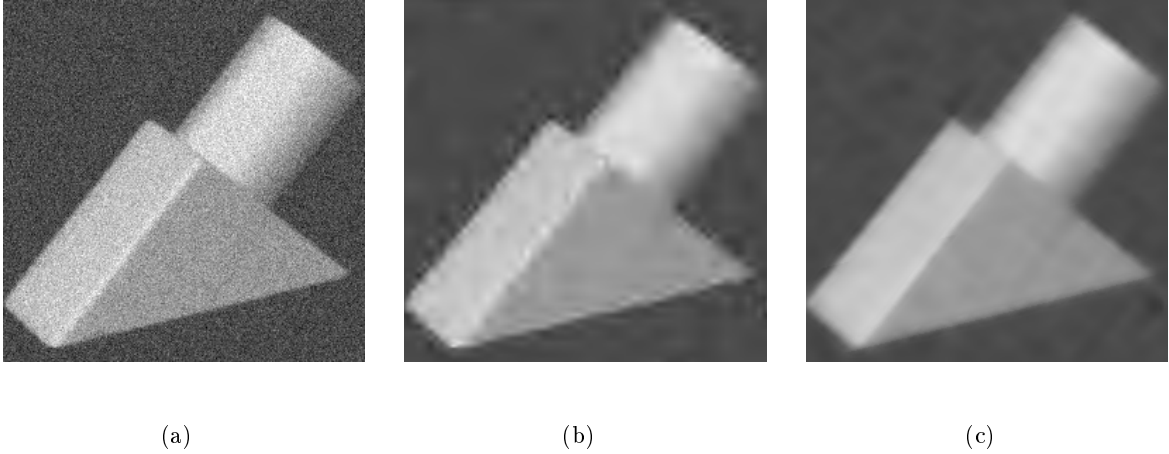

(a)   (b)   (c)

Fig. 15. Comparison of denoising on a real image of size 256×256 with linear structure. (a) Noisy Image; SNR = 9.52 dB. (b) Denoise using 2-D DWT; SNR = 19.78 dB. (c) Denoise using FRIT and Wiener filter; SNR = 21.07 dB.

VIII. Conclusion and Discussion

Several discrete transforms based on the ridgelet idea were presented. Table II summarizes those transforms. Experimental results indicate that the FRIT offers an efficient representation for images that are smooth away from straight edges.

Transform	Image size	Critical sampled	Orthogonal	Comments
FRIT	$p \times p$	Yes	Yes	Basic scheme
Folded FRIT	$n \times n \ (2n - 1 = p)$	Yes	No	Reduce border artifacts
Multiscale FRIT	$n \times n \ (n = p^J q)$	Yes	Yes	Block bases
Galois FRIT	$n \times n \ (n = p^r)$	Yes	Yes	Unnatural lines
Ridgelet frames	any size	No	No	Ill-conditioned

TABLE II

Summary of presented discrete ridgelet transforms.

The discrete ridgelet transforms presented in this paper can be used as a building block in obtaining new schemes which can deal efficiently with natural images. Those images are typically piecewise smooth away from singularities along boundaries that are also smooth curves. Since

27

ridgelets are specially adapted only to straight singularities, a more practical transform would first utilize a quad-tree division of images into localized blocks where edges look straight and then apply the discrete ridgelet transform to each block.

However, a drawback with the ridgelet transform is that its basis functions generally have larger support (which is due to the long length of the ridge function) than those of the wavelet transform. As a result, we observed that the ridgelet transform makes rapid progress in reconstructing the image after the first few coefficients but then slows down significantly; many additional ridgelet coefficients are then needed to correct the mismatch with the image from the first few coefficients. This behavior is also observed for many non-local bases such as the Fourier transform.

One suggestion to obtain more localized ridgelet transform is to redefine its basis functions into some multiscale structure where both the length and the width of these functions are related to the scale. This is reminiscent of the *curvelet* construction [32].

Our construction for digital ridgelets is based on the Radon transform. While this link holds in the continuous space, it has shortcomings in the discrete space. A more "correct" discretization of ridgelet transform would be the one in which the number of directions double at every finer scale. However, in the Radon-based approach, there is a same number of directions at each scale. Thus one may look for a different sampling scheme. We hope to address these issues in a forthcoming paper.

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. We will show here one way of achieving this for $p = n + 1$, where $n = 2^J$. The method can be generalized for other prime numbers.

The idea is to extend the orthonormal wavelet basis vectors from a J-level and periodic wavelet transform of length n to vectors of length p while preserving the zero sum property. Denote $\{v_m : 0 \leq m < n\}$ to be the basis vectors of a such length n wavelet transform where v_0 corresponds to the single scaling coefficient or the mean value, thus all other vectors must have zero

sum. We then define p vectors in $\mathbb{R}^p$ as

$$
\begin{aligned}
\boldsymbol{w}_0 &= (1, 1, \ldots, 1) \, / \, p^{1/2}, \\
\boldsymbol{w}_m &= (\boldsymbol{v}_m, 0), \quad m = 1, 2, \ldots, n - 1, \\
\boldsymbol{w}_{p-1} &= (1, 1, \ldots, 1, -(p-1)) \, / \, [p(p-1)]^{1/2}.
\end{aligned}
$$

It is easy to show that $\{\boldsymbol{w}_m : 0 \le m < p\}$ is an orthonormal basis of $\mathbb{R}^p$ and satisfies *Condition Z*. For an input vector of length p: $\boldsymbol{x} = (x_0, x_1, \ldots, x_{p-1})$, the transform coefficients correspond to $\boldsymbol{w}_m$, $1 \le m \le n - 1$, and can be computed efficiently via the usual DWT of dyadic length n with J-levels on the vector $\boldsymbol{x}' = (x_0, x_1, \ldots, x_{n-1})$. 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] K. R. Rao and P. Yip, *Discrete Cosine Transform: Algorithms, Advantages, Applications*, Academic Press, 1990.

[4] D. Donoho and I. Johnstone, "Ideal spatial adaptation via wavelet shrinkage," *Biometrika*, pp. 425–455, December 1994.

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

[6] E. J. Candès, *Ridgelets: theory and applications*, Ph.D. thesis, Department of Statistics, Stanford University, 1998.

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

[8] S. R. Deans, *The Radon transform and some of its applications*, John Wiley & Sons, 1983.

[9] 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.

[10] J. L. Starck, E. J. Candès, and D.L. Donoho, "The curvelet transform for image denoising," *IEEE Trans. Image Proc.*, 2000, submitted.

[11] S. Mallat and W.L. Hwang, "Singularity detection and processing with wavelets," *IEEE Trans. Inform. Th.*, vol. 38, pp. 617–643, Mar. 1992.

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

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

[14] 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.

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

[16] 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.

[17] 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.

[18] 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.

[19] 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.

[20] 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.

[21] 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.

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

[23] V. K. Goyal and J. Kovačević, "Quantized frame expansions with erasures," *Applied and Computational Harmonic Analysis*, 2001, to appear.

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

[25] C. M. Brislawn, "Classification of nonexpansive symmetric extension transforms for multirate filter banks," *Applied and Computational Harmonic Analysis*, vol. 3, pp. 337–357, 1996.

[26] C. Herley and M. Vetterli, "Wavelets and recursive filter banks," *IEEE Trans. Signal Proc.*, vol. 41, no. 8, pp. 2536–2556, August 1993.

[27] G. Caire, R. L. Grossman, and H. V. Poor, "Wavelet transforms associated with finite cyclic groups," *IEEE Trans. Inform. Th.*, vol. 39, pp. 1157–1166, 1993.

[28] K. Flornes, A. Grossmann, M. Holschneider, and B. Torrésani, "Wavelets over discrete fields," *Applied and Computational Harmonic Analysis*, vol. 1, pp. 137–146, 1994.

[29] Y. Censor, "Finite series-expansion reconstruction methods," *Proc. of IEEE*, vol. 71, no. 3, pp. 409–419, Mar 1983.

[30] S. G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Transactions on Signal Processing, Special Issue on Wavelets and Signal Processing*, vol. 41, no. 12, pp. 3397–3415, December 1993.

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

[32] 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.