Proc. of SPIE Conf. on Wavelet Applications in Signal and Image Processing VIII, San Diego, 2000

Image Denoising using Orthonormal Finite Ridgelet Transform

Minh N. Do* and Martin Vetterli*†

*Laboratory for Audio-Visual Communications (LCAV)
Swiss Federal Institute of Technology Lausanne (EPFL)
CH-1015 Lausanne, Switzerland

†Department of Electrical Engineering and Computer Science
University of California Berkeley, Berkeley CA 94720, USA

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

ABSTRACT

This paper addresses the image denoising problem using a newly proposed digital image transform: the finite *ridgelet transform (FRIT). The transform is invertible, non-redundant and achieved via fast algorithms. Furthermore this transform can be designed to be* orthonormal *thus indicating its potential in many other image processing applications. We then propose various improvements on the initial design of the FRIT in order to make it to have better energy compaction and to reduce the border effect. Experimental results show that the new transform outperforms wavelets in denoising images with linear discontinuities.*

1. INTRODUCTION

State-of-the-art image denoising algorithms are based on thresholding of wavelet coefficients.[1,2] Despite their simplicity, these methods work very effectively. This is based on the fact that in the wavelet domain, maximum image information is packed into a small number of coefficients – precisely at the points of singularities in the spatial domain.[3] However, since wavelets fail to represent efficiently singularities along lines or curves, wavelet-based techniques fail to explore the geometrical singularity of edges in images. Hence, new denoising schemes which are based on true 2-D transforms are expected to improve the performance over traditional wavelet-based approaches.

In order to overcome the weakness of wavelets in higher dimensions, Candes and Donoho[4,5] recently proposed a new system of representations named *ridgelets* which they showed to deal effectively with *linelike* phenomena in 2-D. Much of the work in ridgelets has been concentrated in the *continuous* $\mathbb{R}^2$ space. For practical applications, however *discrete-time* (space) implementations of the ridgelet transform and its reconstruction on a finite $\mathbb{Z}^2$ grid are challenging problems. Due to the radial nature of ridgelets, straightforward implementations based on discretization of continuous formulae would require interpolation in the polar coordinate and thus the resulting transforms would be either redundant or can not be perfectly reconstructed. The digital ridgelet transform used in[6] that is based on *rectopolar* coordinates has a factor 2 overcompleteness and its inversion requires an iterative algorithm.

In this paper, we propose one implementation of the ridgelet transform for digital images that achieves both nonredundant and invertible requirements. In fact the new transform is shown to be orthonormal. Its applications are demonstrated and studied in the image denoising problem.

There are other denoising approaches that explores the geometrical regularity of edges by chaining adjacent wavelet coefficients and then thresholding them over those contours.[7] However, the orthonormal finite 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 edges.

2. CONTINUOUS RIDGELET TRANSFORM

We start by briefly review the continuous ridgelet transform (CRT) and show its connection with the continuous wavelet transform. Given an integrable bivariate function $f(\boldsymbol{x})$, the CRT in $\mathbb{R}^2$ can be defined as follows[4,5]

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

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

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

In comparison, the (separable) continuous wavelet transform (CWT) in $\mathbb{R}^2$ can be written as

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

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, $\psi_{a,b}(t) = a^{-1/2}\psi((t-b)/a)$.*

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 brief, those 2-D transforms are related by:

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

The consequence of this is: as wavelet analysis is very effective at representing objects with isolated point singularities, ridgelet analysis can be 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

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

then the ridgelet transform is precisely the application of a 1-D wavelet transform to the slices of the Radon transform,

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

It is interesting to note that if in (6) instead of taking 1-D wavelet transform, the application of 1-D Fourier transform on t would result in the 2-D Fourier transform, or

$$F_f^2(\xi\cos\theta, \xi\sin\theta) = \int_{\mathbb{R}} e^{-j\xi t} RA_f(\theta,t) dt. \tag{7}$$

This is the famous *projection-slice* theorem and is used often in image reconstruction from projection methods.[8]

*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.

3. FINITE RADON TRANSFORM

As suggested from the previous section, a discrete-time ridgelet transform can be obtained using a discrete-time Radon transform. Numerous discretization of the Radon transforms have been devised to approximate the continuous formula.[8] However to our knowledge, none of them were designed to be as invertible transforms for digital images. On the other hand, the *finite*[†] Radon transform theory[9–11] in combinatorics provides an interesting alternative.

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 $Z_p = \{0, 1, \ldots, p-1\}$, where p is a prime number. Note that Z_p is a finite field with modulo p operations.

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 specifically

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

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) and we assume this in the sequel. The factor $\frac{1}{\sqrt{p}}$ is introduced to normalize the l_2-norm.

So 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 ($k = p$ corresponds to infinite slope or vertical lines) and its intercept l. 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. That means

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

Equation (10) explicitly reveal 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, i.e.

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

where $P_{i,j}$ denotes the set of indexes of all the lines that go through the point (i,j). More precisely, using (9) we can write

$$P_{i,j} = \{(k,l) : l = j - ki \pmod p, k \in Z_p\} + \{(p,i)\} \tag{12}$$

From the properties of the finite geometry that every two points lie in exactly one line and there are $p+1$ lines that go through each point, substituting (8) into (11) leads to

$$FBP_r[i,j] = \frac{1}{p}\left(\sum_{(i',j')\in Z_p^2} f[i',j'] + pf[i,j] \right) = f[i,j].$$

[†]That means transform for *finite* length signals.

So the back-projection operator defined in (11) indeed computes the inverse FRAT. It is worth to note that the transform matrices for the operators $FRAT$ and FBP are transposed of each other. In other word, for the subspace of zero mean functions defined on $\mathbb{Z}_p^2$, the FRAT is a *tight frame*.

Both the forward and inverse of the FRAT requires p^3 additions and p^2 multiplications. In,[10] Matúš and Flusser described a fast implementation of the FRAT in that for each projection we need to pass all pixels of the original image once and use p histogrammers: one for each FRAT coefficient of that projection. On a Sun Ultra 5 computer, both forward and inverse FRAT take least than 1 second for an image of size 257×257.

For any set A, write δ_A for the characteristic function of A. Then we can write the basis functions for the FRAT as $\{p^{-1/2}\delta_{L_{k,l}} : \ 0 \le k \le p,\ 0 \le l < p\}$. The previously mentioned properties of lines in the finite geometry Z_p^2 yield

$$\langle \delta_{L_{k,l}}, \delta_{L_{k',l'}} \rangle = \begin{cases} p & \text{if } k = k', l = l' \\ 0 & \text{if } k = k', l \neq l' \\ 1 & \text{if } k \neq k' \end{cases} \tag{13}$$

Hence the minimum angle between any basis functions of the FRAT is: $\min_{k,l,k',l'} \angle(\delta_{L_{k,l}}, \delta_{L_{k',l'}}) = \cos^{-1}(1/p)$, which approaches the right angle for large p. So we can say that the finite Radon basis is "almost" orthogonal.

4. ORTHOGONAL 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 vector, also called a *projection*, $(r_k[0], r_k[1], \ldots, r_k[p-1])$, of Radon coefficients where the direction k is fixed.[‡] The result can be called as *finite ridgelet transform* (FRIT). Due to periodicity of the FRAT coefficients for each direction, periodic wavelet transforms are chosen and assumed in the sequel.

Note 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 filterbank with J states where G_0 and G_1 are low and high pass synthesis filters, respectively. Then the family of functions $\{g_1^{(j)}[n-2^j m], g_0^{(J)}[n-2^J m]\}$, $j = 1, \ldots, J$, and $n, m \in \mathbb{Z}$ is the orthogonal basis of the discrete-time wavelet series.[12] Here $G^{(j)}$ are the equivalent filters at level j. The basis functions from $g_0^{(J)}$ are called the scaling functions where the others are called wavelet functions. Normally, the filter G_1 is designed to satisfy the high pass condition, $G_1(z)|_{z=1} = 0$. Therefore each wavelet basis function has zero sum.

For a general setting, let's 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

$$\{\boldsymbol{w}_m^{(k)} : \ m \in Z_p\}, \qquad 0 \le k \le p.$$

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

LEMMA 4.1 (CONDITION Z). *Suppose that $\{\boldsymbol{w}_m, \in Z_p\}$ 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 sum.*

Proof: Denote $\mathbf{1} = (1, 1, \ldots, 1)$, $\mathbf{1} \in \mathbb{R}^p$. 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, let's assume that each basis function $\boldsymbol{w}_m$, $1 \le m \le p-1$ 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 $S_0 = \{c\mathbf{1} : c \in \mathbb{R}\}$. We have $\langle c\mathbf{1}, \boldsymbol{w}_i \rangle = c \sum_l w_m[l] = 0$, $\forall m = 1, \ldots, p-1$,

[‡] If p is not dyadic, a special border handling is required.

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 our definition, the FRIT is can now be written as

$$\begin{aligned} FRIT_f[k,m] &= \langle FRAT_f[k,.], w_m^{(k)}[.]\rangle \\ &= \sum_{l\in Z_p} w_m^{(k)}[l]\langle f, p^{-1/2}\delta_{L_{k,l}}\rangle. \end{aligned} \tag{14}$$

Hence we can write the basis functions for the FRIT as follows

$$\rho_{k,m} = \frac{1}{\sqrt{p}} \sum_{l\in Z_p} w_m^{(k)}[l]\ \delta_{L_{k,l}}. \tag{15}$$

Let's consider the inner products between any two FRIT basis functions

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

Using (13), 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 $\{\boldsymbol{w}_m^{(k)} :\ m \in Z_p\}$. Next for the case when the two FRIT basis functions have different directions, $k \neq k'$, again using (13) 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 direction k (see (10)). So together with the assumption that $\boldsymbol{w}_0^{(k)}$ are a constant functions we see that all of the FRIT basis functions $\rho_{k,0}$ 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. Thus we complete the proof of the following result.

THEOREM 4.2. *Given $p+1$ orthonormal bases in $l^2(Z_p)$ (which can be the same): $\{\boldsymbol{w}_m^{(k)},\ m \in Z_p\}$, $0 \le k \le p$, that satisfy the* Condition Z *then*

$$\{\rho_{k,m} :\ 0 \le k \le p,\ 0 < m < p\} \cup \{\rho_0 = 1/p\}$$

is an orthonormal basis in $l^2(Z_p^2)$, where $\rho_{k,m}$ are defined in (15).

Remarks

1. The 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 all the DWT's are taken to the maximum number of levels then it happens that all of the remaining scaling coefficients at different projections are the same, hence we can drop all but one of them. The result is the orthonormal FRIT.

2. We prove the above result for the general setting where different transforms can be applied on different FRAT projections. This allows us to use adaptive bases, such as wavelet packets, on each projection independently. Furthermore, due to the "wrap around" effect of the FRAT, its projections could have strong periodicity components so for some projections one could use a Fourier-type transform like DCT. Note[10] that if we apply the 1-D Fourier transform on all of the FRAT projections then we would obtain the 2-D Fourier transform.

3. All of these results also hold for any other affine planes rather than Z_p^2. Especially using Galois planes constructed by the finite field $GF(p^r)$ we can build orthogonal FRIT's for images of size $n \times n$ where n is a power of a prime, $n = p^r$.

Figure 1 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.

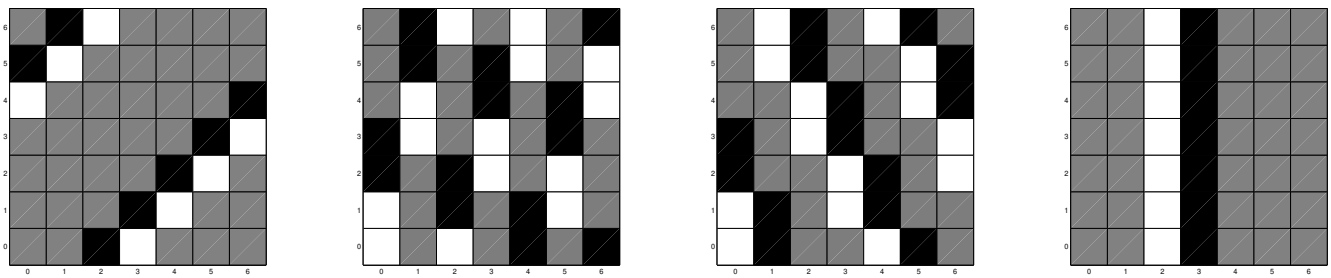


Figure 1. Examples of 7×7 FRIT basis functions using Haar wavelets. Black, grey and white pixels corresponding respectively to negative, zero and positive values. Notice the "wrap-around" effect due to periodization of FRAT.

5. OPTIMAL ORDER OF THE FINITE RADON COEFFICIENTS

The FRAT described in Section 3 use (9) as a convenient way to specify finite lines on the Z_p^2 lattice via two parameters k and l. However it is neither the unique nor the best way for our purpose. In fact, the most general way of defining lines on Z_p^2 plane is

$$L'_{a,b,t} = \{(i,j) \in Z_p^2 : ai + bj - t = 0 \pmod p\}, \tag{16}$$

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 line defined as in (16), (a,b) is its finite normal vector and t is its translation parameter. Note that 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 (16), consider the line equation

$$ai + bj - t = 0. \tag{17}$$

If $b \neq 0$ then (17) $\Leftrightarrow j = -b^{-1}ai + b^{-1}t$, where b^{-1} denotes the multiplicative inverse of b in the finite field Z_p, i.e. $bb^{-1} = 1 \pmod p$. For the case $b = 0$, it is necessary that $a \neq 0$, thus (17) $\Leftrightarrow i = a^{-1}t$.

So we have the following equivalence between two specifications for finite lines:

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

Furthermore, it is easy to verify that for $c \in Z_p, c \neq 0$ then $\{cl : l \in Z_p\} = Z_p$. So for a fixed normal vector (a, b), the set of lines $\{L'_{a,b,t} : t \in Z_p\}$ is precisely the set of p parallel lines with same slope k defined previously, $\{L_{k,l} : l \in Z_p\}$, where $k = -b^{-1}a$ for $b \neq 0$ and $k = p$, for $b = 0$. In other words, the set of lines with normal vector (a, b) is exactly the same with the set of lines with normal vector (na, nb), $\forall n = 1, 2, \ldots, p-1$.

Now the FRAT using the new line specification (16) can be written as

$$r_{a,b}[t] = FRAT_f[a, b, t] = \frac{1}{\sqrt{p}} \sum_{(i,j) \in L'_{a,b,t}} f[i, j]. \tag{18}$$

From the discussion above we see that the vector $(r_{a,b}[0], r_{a,b}[1], \ldots, r_{a,b}[p-1])$ is simply a reorder of a FRAT projection vector $(r_k[0], r_k[1], \ldots, r_k[p-1])$ defined previously. The standard FRAT described in Section 3 uses the set of $(p+1)$ normal vectors $\boldsymbol{u}_k$, $0 \leq k \leq p$, where

$$\begin{aligned} \boldsymbol{u}_k &= (-k, 1) \qquad \text{for } k = 0, 1, \ldots, p-1, \text{ and} \\ \boldsymbol{u}_p &= (1, 0). \end{aligned} \tag{19}$$

In order to provide a complete representation, the FRAT can be defined as in (18) with any set of $(p+1)$ normal vectors $\{(a_k, b_k), 0 \leq k \leq p\}$ which covers all $p+1$ distinct directions (slopes) in the finite geometry $\mathbb{Z}_p^2$. There are $p-1$ such choices for each slope k: $(a_k, b_k) = n\boldsymbol{u}_k$, $1 \leq n \leq p-1$.

The ordering of FRAT coefficient is important for FRIT since we apply 1-D DWT on the FRAT projections. This order depends on the choice of the normal vector (a_k, b_k). So which is the optimal set of $(p+1)$ 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.[10]

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

$$F[u, v] = \frac{1}{p} \sum_{(i,j) \in Z_p^2} f[i, j] W_p^{ui+vj}, \tag{20}$$

and for FRAT projections on Z_p as

$$R_{a,b}[w] = \frac{1}{\sqrt{p}} \sum_{t \in Z_p} r_{a,b}[t] W_p^{wt}. \tag{21}$$

THEOREM 5.1 (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):*

$$R_{a,b}[w] = F[aw, bw]. \tag{22}$$

Proof: Substituting (18) into (21) and using the fact that the set of p parallel lines $\{L'_{a,b,t} : t \in Z_p\}$ provides a complete cover of the plane Z_p^2, we obtain

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

□

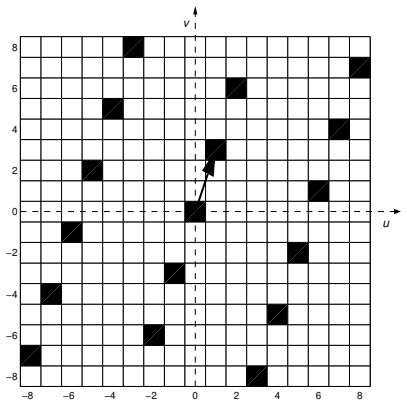


Figure 2. Example of a discrete Fourier slice with the best normal vector (the solid arrow) for that direction. In this example, $p = 17$, the slope of the direction is $k = 11$ and the best normal vector is $(1, 3)$.

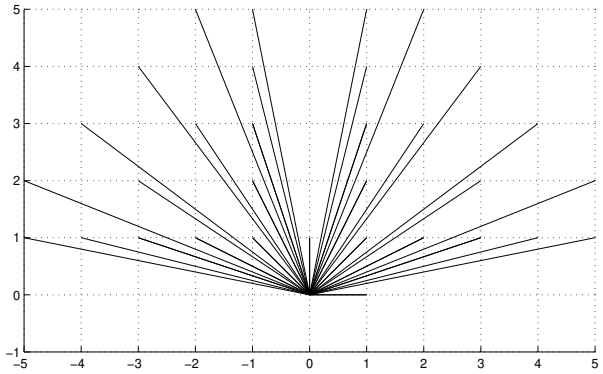


Figure 3. Optimal set of normal vectors for FRAT of size $p = 31$.

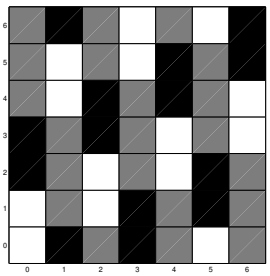
From (22) we see clearly the role of the normal vectors (a_k, b_k) in FRAT. As in the case for the projections, the chosen normal vector (a, b) for a particular direction controls the order of the coefficients in the corresponding Fourier slices. For the kind of images that we are interested in – i.e. natural images – most of the energy is concentrated in the low frequency. On the other hand, in the wavelet transform the frequency division is also concentrated in the low frequency. Thus in order to ensure that the FRAT projections to be a smooth, low frequency dominated signals, the points (a, b) should be chosen to be as "close" to the origin of the Fourier plane as possible.

Furthermore, the choice of the closest point (a, b) to the origin directs the represented direction of a given Fourier slice is the one that has least "wrap around" – parts of the slice that do not cross the origin – in the the frequency plane. Figure 2 illustrates this by showing an example of discrete Fourier slice. The normal vector for this particular direction of FRAT can be chosen as a vector from the origin to any other points on the Fourier slice. The best normal vector is selected as the closest point to the origin.

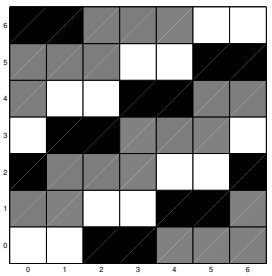
Formally, we define the set of $p + 1$ optimal normal vectors $\{(a_k^\star, b_k^\star),\ k = 0, 1, \ldots, p\}$ as follows:

$$(a_k^\star, b_k^\star) = \arg \min_{\substack{(a_k, b_k) \in \{n\mathbf{u}_k, 1 \le n \le p-1\} \\ \text{s.t. } c_p(b_k) \ge 0}} ||(\mathcal{C}_p(a_k), \mathcal{C}_p(b_k))|| \qquad k = 0, 1, \ldots, p. \tag{23}$$

Here $\mathcal{C}_p(x)$ is the centralized function of period p: $\mathcal{C}_p(x) = x - p.\text{round}(x/p)$. The constraint $\mathcal{C}_p(b_k) \ge 0$ is imposed in order to remove the ambiguity in deciding between (a, b) and $(-a, -b)$. As a result, the optimal normal vectors are restricted to have angles in $[0, \pi)$. We use norm-2 for solving (23). Figure 3 shows an example result of the optimal set of normal vectors. In comparison with the usual set of normal vectors $\{\boldsymbol{u}_k, 0 \le k \le p\}$, the new set $\{(a_k^\star, b_k^\star), 0 \le k \le p\}$ provides a much more equal coverage in angular directions.



(a) With the standard ordering of FRAT



(b) With the optimal ordering of FRAT

Figure 4. Example of the same 7×7 FRIT basis function that uses two different ordering of FRAT coefficients.

After computing the normal vector set $\{(a_k^\star, b_k^\star), 0 \le k \le p\}$, the FRAT and its inverse are processed with the same fast algorithms as above. For a given p, solving (23) requires $O(p^2)$ operations. Therefore it is negligible compared with the transforms themselves.

The effect of the new ordering of FRAT coefficients on the FRIT is shown in Figure 4. As can be seen, with the optimal ordering, the adjacent FRAT basis functions in a direction are closer together hence the resulting FRIT basis functions resemble much more the discrete linear singularities.

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

Assuming that p is a prime number $p > 2$, then p is odd and can be written as $p = 2n - 1$. Consider an input image of size $n \times n$, $f[i, j]$, $0 \le i, j < n$. This image is folded with respect to the lines $i = 0$ and $j = 0$ to produce a $p \times p$ image $\check{f}[i, j]$, $-n < i, j < n$, in which

$$\check{f}[i, j] = f[|i|, |j|]. \tag{24}$$

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 symmetric properties so that the original image can be perfectly reconstructed by keeping only about 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 (24), we obtain

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

Theorem 5.1 shows that the FRAT $\check{r}_{a,b}[t]$, $-n < t < n$, can be computed from the inverse 1-D DFT

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

where $\check{R}_{a,b}[w] = \check{F}[aw, bw]$.

The symmetry of $\check{F}[u, v]$ thus yields

$$\check{R}_{a,b}[w] = \check{R}_{a,b}[|w|] \quad \text{and} \tag{25}$$

$$\check{R}_{a,b}[w] = \check{R}_{|a|,|b|}[w]. \tag{26}$$

From (25) 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 (26) reveals the duplications among those projections. In fact, using the set of optimal normal vectors in (23), 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. So the FRAT of $\check{f}[i, j]$ is completely specified by the following $n(n + 1)$ coefficients:

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

Furthermore, as in (10) the redundancy of those coefficients is explicitly specified by

$$\check{r}_{a_k^\star, b_k^\star}[0] + 2 \sum_{t=1}^{n-1} \check{r}_{a_k^\star, b_k^\star}[t] = 0 \tag{27}$$

where again we assume that the mean is subtracted from the image $\check{f}[i, j]$ previous to taking the FRAT. Therefore we can extract exactly n^2 independent folded FRAT coefficients so that the original image $f[i, j]$ can be perfectly reconstructed.

However orthogonality might be lost for the folded FRIT (resulting from applying 1-D DWT on $n+1$ projections of the folded FRAT). This is due to the fact that the basis functions from a same direction of the folded FRAT could have overlap. On the other hand, if we loosen up the orthogonality constraint, then by construction the folded FRAT projections $(\check{r}_{a_k^\star, b_k^\star}[t], 0 \leq t < n)$ are symmetric with respect to $t = 0$ and $t = n - 1/2$. This allows the use of folded wavelet transform with biorthogonal symmetric wavelets.[13]

7. IMAGE DENOISING EXPERIMENTS

The intuition behind the FRIT-based method in image denoising is that typical linear singularities of image are represented by a few coefficients while randomly located noisy singularities are unlikely to produce significant coefficients. This is unlike the wavelet transform where both noisy pixels and edges can produce large amplitude wavelet coefficients. Therefore simple thresholding FRIT coefficients scheme 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[1] to the FRIT coefficients. More specifically, our denoising algorithm consists of the following steps:

1. Apply the FRIT to the noisy image.
2. Hard-thresholding FRIT coefficients with the universal threshold $T = \sigma\sqrt{2logN}$ where $N = p^2$ pixels.
3. Inverse FRIT of the thresholded coefficients.

   For a image which is smooth away from linear singularities, edges are visually well restored after Step 3. However due to the periodization 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.

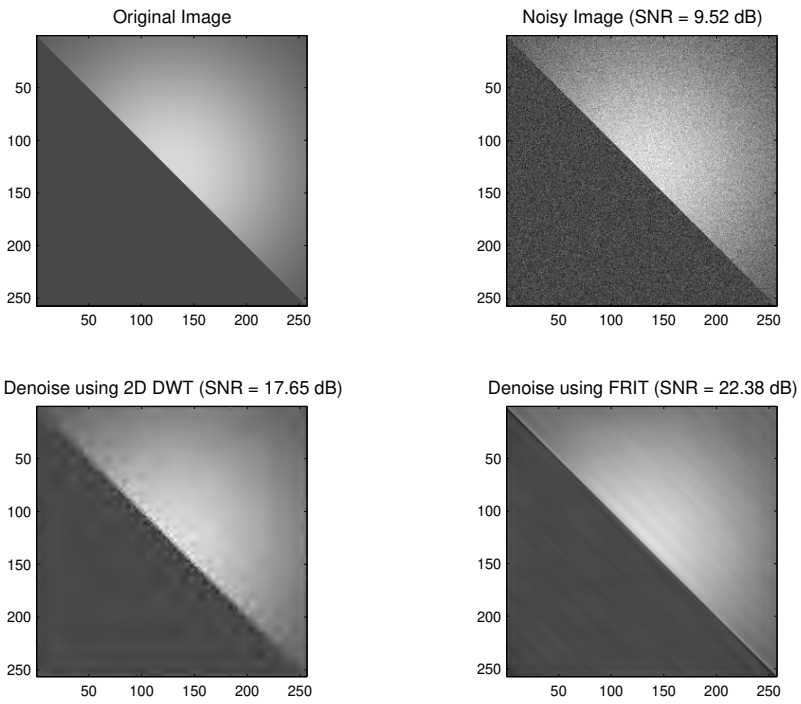


Figure 5. Comparison of denoising on a half-plane truncated Gaussian image of size 257×257.

The above denoising algorithm is compared against the 2-D wavelet hard-thresholding method using the same threshold value.

Using a similar image to the one studied in,[14] our first test image is the half-plane truncated Gaussian function, $f(x_1, x_2) = 1_{\{x_1+x_2>0\}} e^{-x_1^2-x_2^2}$, $\boldsymbol{x} \in \mathbb{R}^2$. Figure 5 displays the comparison of denoising results on this image using the 2-D DWT and the FRIT. The FRIT clearly outperforms the 2-D DWT; especially it correctly "pick ups" the edge.

Our second test is a real image of an object with linear structure. Figure 6 displays the result. Again we can see that in contrast with the DWT, edges are restored very well with the FRIT.

8. DISCUSSION AND FURTHER WORK

A new finite orthonormal image transform based on the ridgelet idea is presented. The FRIT is shown to represent effectively images with linear discontinuities – i.e. images with straight edges. Experimental results indicate that FRIT has potential in restoring images that are smooth away from straight edges. Furthermore, by applying on suitable block sizes, this transform is expected to work well also for images with smooth edges

However, the FRIT has a drawback which is due to its annoying "wrap-around" effect. In non-linear approximation, large FRIT coefficients that are responses to a linear edge also create some periodic lines in the reconstructed images. We are working on a new construction of the discrete Radon transform which is free from this periodization.

Another related problem with the ridgelet transform is that its basis functions generally have larger support – i.e. the long length of the ridge function – than those of the wavelet transform. Thus a discontinuity in the input image can create many more non-zero ridgelet coefficients. As a result, the ridgelet transform makes rapid progress in reconstructing the image after the first few coefficients but then slows down significantly. Further ridgelet coefficients are then added in to correct the minor mismatch with the image from the first few coefficients. This behavior is well-known for any 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 decay with the scale. This somewhat coincides with the *curvelet* construction.[15]

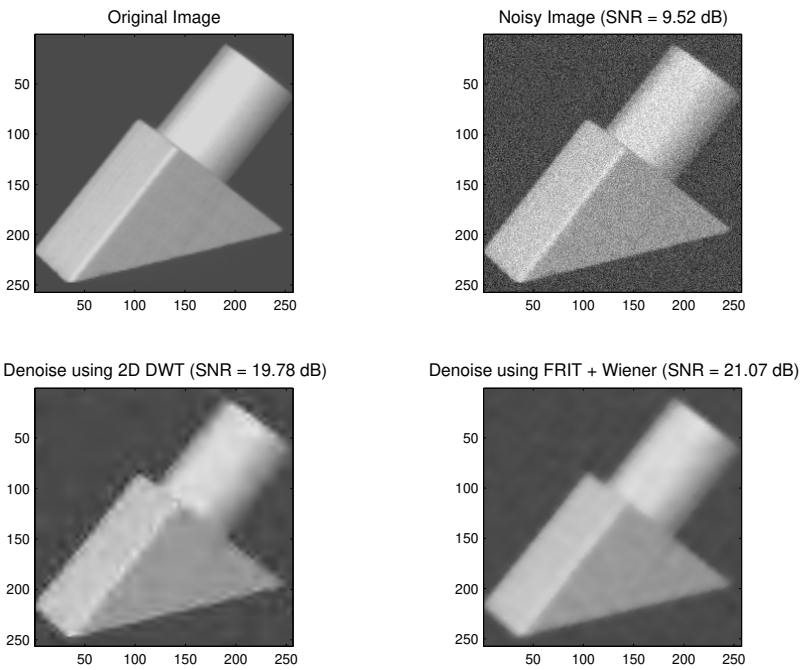


Figure 6. Comparison of denoising on an image of a real object with linear structure.

Acknowledgment

This work was supported by a Department of Communication Systems, EPFL PhD Fellowship and the Swiss National Science Foundation grant number 21-52439.97.

REFERENCES

1. D. Donoho and I. Johnstone, "Ideal spatial adaptation via wavelet shrinkage," *Biometrika* , pp. 425–455, December 1994.
2. S. Mallat, *A Wavelet Tour of Signal Processing*, Academic Press, 2nd ed., 1999.
3. D. L. Donoho, M. Vetterli, R. A. DeVore, and I. Daubechies, "Data compression and harmonic analysis," *IEEE Trans. Info. Theory* **44**, pp. 2435–2476, October 1998.
4. E. Candès, *Ridgelets: theory and applications*. PhD thesis, Department of Statistics, Stanford University, 1998.
5. E. Candès and D. L. Donoho, "Ridgelets: a key to higher-dimensional intermittency?," *Phil. Trans. R. Soc. Lond. A.* , pp. 2495–2509, 1999.
6. D. L. Donoho and M. R. Duncan, "Digital curvelet transform: strategy, implementation and experiments," tech. rep., Department of Statistics, Stanford University, 1999.
7. S. Mallat and W. Hwang, "Singularity detection and processing with wavelets," *IEEE Trans. Info. Theory* **38**, pp. 617–643, Mar. 1992.
8. G. T. Herman, *Image Reconstruction from Projections: The Fundamentals of Computerized Tomography*, Academic Press, 1980.
9. E. D. Bolker, "The finite Radon transform," in *Integral Geometry (Contemporary Mathematics, Vol. 63)*, S. H. R. L. Bryant, V. Guillemin and R. O. W. Jr., eds., pp. 27–50, 1987.
10. F. Matúš and J. Flusser, "Image representation via a finite Radon transform," *IEEE Trans. Pattern Anal. Machine Intell.* **15**, pp. 996–1006, Oct 1993.
11. 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.
12. M. Vetterli and J. Kovačević, *Wavelets and Subband Coding*, Prentice-Hall, Englewood Cliffs, NJ, 1995.
13. C. M. Brislawn, "Classification of nonexpansive symmetric extension transforms for multirate filter banks," *Applied and computational harmonic analysis* **3**, pp. 337–357, 1996.
14. E. Candès, "Ridgelets: Estimating with ridge functions," tech. rep., Department of Statistics, Stanford University, 1999.
15. E. Candès and D. L. Donoho, "Curvelets – a suprisingly effective nonadaptive representation for objects with edges," in *Proc. of Fifth Intl. Conf. on Mathematical Methods for Curves and Surfaces*, L. L. Schumaker *et al.*, eds., 2000. to appear.