Sparse directional image representations using the discrete shearlet transform

Glenn Easley [a], Demetrio Labate [b,*,1], Wang-Q Lim [c]

[a] *System Planning Corporation, Arlington, VA 22209, USA*
[b] *Department of Mathematics, North Carolina State University, Campus Box 8205, Raleigh, NC 27695, USA*
[c] *Department of Mathematics, Lehigh University, Bethlehem, PA 18015, USA*

Abstract

In spite of their remarkable success in signal processing applications, it is now widely acknowledged that traditional wavelets are not very effective in dealing multidimensional signals containing distributed discontinuities such as edges. To overcome this limitation, one has to use basis elements with much higher directional sensitivity and of various shapes, to be able to capture the intrinsic geometrical features of multidimensional phenomena. This paper introduces a new discrete multiscale directional representation called the discrete shearlet transform. This approach, which is based on the shearlet transform, combines the power of multiscale methods with a unique ability to capture the geometry of multidimensional data and is optimally efficient in representing images containing edges. We describe two different methods of implementing the shearlet transform. The numerical experiments presented in this paper demonstrate that the discrete shearlet transform is very competitive in denoising applications both in terms of performance and computational efficiency.
Published by Elsevier Inc.

Keywords: Curvelets; Denoising; Image processing; Shearlets; Sparse representation; Wavelets

1. Introduction

One of the most useful features of wavelets is their ability to efficiently approximate signals containing pointwise singularities. Consider a one-dimensional signal $s(t)$ which is smooth away from point discontinuities. If $s(t)$ is approximated using the best M-term wavelet expansion, then the rate of decay of the approximation error, as a function of M, is optimal. In particular, it is significantly better than the corresponding Fourier approximation error [14,32].

However, it is now widely acknowledged that traditional wavelet methods do not perform as well with multidimensional data. Indeed wavelets are very efficient in dealing with pointwise singularities *only*. In higher dimensions, other types of singularities are usually present or even dominant, and wavelets are unable to handle them very efficiently.

Images, for example, typically contain sharp transitions such as edges, and these interact extensively with the elements of the wavelet basis. As a result, "many" terms in the wavelet representation are needed to accurately represent these objects. In order to overcome this limitation of traditional wavelets, one has to increase their directional sensitivity and a variety of methods for addressing this task have been proposed in recent years. They include several schemes of "directional wavelets" (such as [1,3]), *contourlets* [12,31], *complex wavelets* [25,34], *brushlets* [9], *ridgelets* [6], *curvelets* [7], *bandelets* [29] and *shearlets* [18,20], introduced by the authors and their collaborators.

To make this discussion more rigorous, it will be useful to examine this problem from the point of view of approximation theory. If $\mathcal{F} = \{\psi_\mu : \mu \in I\}$ is a basis or, more generally, a tight frame for $L^2(\mathbb{R}^2)$, then an image f can be (non-linearly) approximated by the partial sums

$$f_M = \sum_{\mu \in I_M} \langle f, \psi_\mu \rangle \psi_\mu,$$

where I_M is the index set of the M largest inner products $|\langle f, \psi_\mu \rangle|$. The resulting approximation error is

$$\varepsilon_M = \|f - f_M\|^2 = \sum_{\mu \notin I_M} |\langle f, \psi_\mu \rangle|^2,$$

and this quantity approaches asymptotically zero as M increases. For many signal processing applications, the goal is to design the representation system $\mathcal{F}$ that achieves the best asymptotic decay rate for this error. For example, for compression applications it has been shown that the distortion rate is proportional ε_M [17]. Similarly, the efficiency of noise removal algorithms that use thresholding estimators have been shown to depend upon ε_M [15].

Let C^2 be the space of functions that are twice continuously differentiable. If the image f is C^2, then the approximation f_M obtained from the M largest wavelet coefficients satisfies

$$\varepsilon_M \leqslant C M^{-2}.$$

However, images typically contain edges. If f is C^2 everywhere away from edge curves that are piecewise C^2, then the discontinuity creates many wavelet coefficients of large amplitude. As a result (see [32]), the asymptotic approximation error obtained using wavelets only decays as

$$\varepsilon_M \leqslant C M^{-1}.$$

This is better than Fourier approximations (in which case the error decays as $M^{-1/2}$), but far from the theoretical optimal approximation, where ε_M decays as M^{-2} [13].

This shows that one can improve upon the wavelet representation by appropriately exploiting the geometric regularity of the edges. Indeed, Candès and Donoho have recently introduced the *curvelet* representation, a tight frame of elongated oscillatory functions at various scales, that produce an essentially optimal approximation rate [7]. Namely, it satisfies

$$\varepsilon_M \leqslant C (\log M)^3 M^{-2}. \tag{1.1}$$

However, curvelets are not generated from the action of a finite family of operators on a single function, as is the case with wavelets. This means their construction is not associated with a multiresolution analysis. This and other issues make the discrete implementation of curvelets very challenging as is evident by the fact that two different implementations of it have been suggested by the originators (see [35] and [5]). In an attempt to provide a better discrete implementation of the curvelets, the *contourlet representation* has been recently introduced [11,31,33]. This is a discrete time-domain construction, which is designed to achieve essentially the same frequency tiling as the curvelet representation (observe however that the contourlets are not a 'discretization' of curvelets).

The authors of this paper and their collaborators have recently introduced the *shearlet representation* [18–22], which yields the same optimal approximation properties (1.1). This new representation is based on a simple and rigorous mathematical framework which not only provides a more flexible theoretical tool for the geometric representation of multidimensional data, but is also more natural for implementation. In addition, the shearlet approach can be associated to a multiresolution analysis [22,27]. In this paper, we will develop discrete implementations of the *shearlet transform* to obtain the discrete shearlet transform. We will show that the mathematical framework of the shearlet transform allows us to develop a simple and faithful transition from the continuous to the discrete representation. It

will become clear from our constructions that the shearlet approach can be viewed as a simplifying theoretical justification for the contourlet transform. The shearlet transform, however, offers a much more flexible approach, and allows one to develop a variety of alternative implementations, with complete control over the mathematical properties of the transform, and features that can be adapted to specific applications.

The paper is organized as follows. In Section 2 we describe the mathematical theory of shearlets and its connection with the theory of affine systems with composite dilations. In Section 3 we introduce the discrete shearlet transform. Finally, in Section 4, we present several demonstrations of the discrete shearlet transform for noise removal applications. Concluding remarks are made in Section 5.

2. Shearlets

The theory of composite wavelets, recently introduced by the authors and their collaborators [20–22], provides an especially effective approach for combining geometry and multiscale analysis by taking advantage of the classical theory of affine systems. In dimension $n = 2$, the *affine systems with composite dilations* are the collections of the form:

$$\mathcal{A}_{AB}(\psi) = \left\{ \psi_{j,\ell,k}(x) = |\det A|^{j/2} \psi \left(B^\ell A^j x - k \right) : j, \ell \in \mathbb{Z}, \ k \in \mathbb{Z}^2 \right\}, \tag{2.2}$$

where $\psi \in L^2(\mathbb{R}^2)$, A, B are 2×2 invertible matrices and $|\det B| = 1$. The elements of this system are called *composite wavelets* if $\mathcal{A}_{AB}(\psi)$ forms a *Parseval frame* (also called *tight frame*) for $L^2(\mathbb{R}^2)$; that is,

$$\sum_{j,\ell,k} \left| \langle f, \psi_{j,\ell,k} \rangle \right|^2 = \|f\|^2,$$

for all $f \in L^2(\mathbb{R}^2)$. In this approach, the dilations matrices A^j are associated with *scale transformations*, while the matrices B^ℓ are associated to area-preserving *geometrical transformations*, such as rotations and shear. This framework allows one to construct Parseval frames whose elements range not only at various scales and locations, like wavelets, but also at various orientations.

In this paper, we will consider a special example of composite wavelets in $L^2(\mathbb{R}^2)$, called shearlets. These are collections of the form (2.2) where $A = A_0$ is the *anisotropic dilation matrix* and $B = B_0$ is the *shear matrix*, which are given by

$$A_0 = \begin{pmatrix} 4 & 0 \\ 0 & 2 \end{pmatrix}, \qquad B_0 = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}.$$

For any $\xi = (\xi_1, \xi_2) \in \hat{\mathbb{R}}^2$, $\xi_1 \neq 0$, let $\psi^{(0)}$ be given by

$$\hat{\psi}^{(0)}(\xi) = \hat{\psi}^{(0)}(\xi_1, \xi_2) = \hat{\psi}_1(\xi_1) \hat{\psi}_2 \left(\frac{\xi_2}{\xi_1} \right), \tag{2.3}$$

where $\hat{\psi}_1, \hat{\psi}_2 \in C^\infty(\hat{\mathbb{R}})$, $\operatorname{supp} \hat{\psi}_1 \subset \left[-\frac{1}{2}, -\frac{1}{16} \right] \cup \left[\frac{1}{16}, \frac{1}{2} \right]$ and $\operatorname{supp} \hat{\psi}_2 \subset [-1, 1]$. This implies that $\hat{\psi}^{(0)}$ is C^∞ and compactly supported with $\operatorname{supp} \hat{\psi}^{(0)} \subset \left[-\frac{1}{2}, \frac{1}{2} \right]^2$. In addition, we assume that

$$\sum_{j \geqslant 0} \left| \hat{\psi}_1 \left(2^{-2j} \omega \right) \right|^2 = 1 \quad \text{for } |\omega| \geqslant \frac{1}{8}, \tag{2.4}$$

and, for each $j \geqslant 0$,

$$\sum_{\ell = -2^j}^{2^j - 1} \left| \hat{\psi}_2 \left(2^j \omega - \ell \right) \right|^2 = 1 \quad \text{for } |\omega| \leqslant 1. \tag{2.5}$$

From the conditions on the support of $\hat{\psi}_1$ and $\hat{\psi}_2$ one can easily observe that the functions $\psi_{j,\ell,k}$ have frequency support:

$$\operatorname{supp} \hat{\psi}_{j,\ell,k}^{(0)} \subset \left\{ (\xi_1, \xi_2) : \xi_1 \in \left[-2^{2j-1}, -2^{2j-4} \right] \cup \left[2^{2j-4}, 2^{2j-1} \right], \ \left| \frac{\xi_2}{\xi_1} + \ell 2^{-j} \right| \leqslant 2^{-j} \right\}.$$
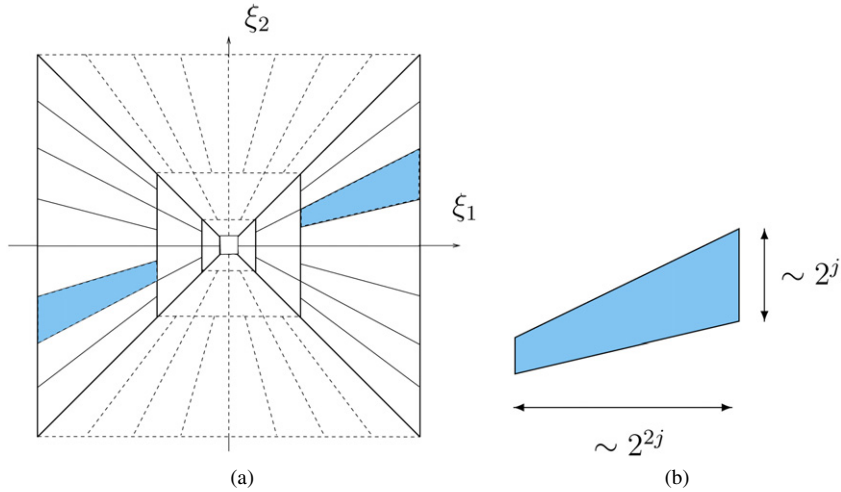
Fig. 1. (a) The tiling of the frequency plane $\hat{\mathbb{R}}^2$ induced by the shearlets. The tiling of $\mathcal{D}_0$ is illustrated in solid line, the tiling of $\mathcal{D}_1$ is in dashed line. (b) The frequency support of a shearlet $\psi_{j,\ell,k}$ satisfies parabolic scaling. The figure shows only the support for $\xi_1 > 0$; the other half of the support, for $\xi_1 < 0$, is symmetrical.

That is, each element $\hat{\psi}_{j,\ell,k}$ is supported on a pair of trapezoids, of approximate size $2^{2j} \times 2^j$, oriented along lines of slope $\ell 2^{-j}$ (see Fig. 1b).

There are several examples of functions ψ_1, ψ_2 satisfying the properties described above (see Appendix A). Equations (2.4) and (2.5) imply that

$$\sum_{j \geqslant 0} \sum_{\ell=-2^j}^{2^j-1} \left|\hat{\psi}^{(0)}\left(\xi A_0^{-j} B_0^{-\ell}\right)\right|^2 = \sum_{j \geqslant 0} \sum_{\ell=-2^j}^{2^j-1} \left|\hat{\psi}_1\left(2^{-2j}\xi_1\right)\right|^2 \left|\hat{\psi}_2\left(2^j \frac{\xi_2}{\xi_1} - \ell\right)\right|^2 = 1$$

for $(\xi_1, \xi_2) \in \mathcal{D}_0$, where $\mathcal{D}_0 = \{(\xi_1, \xi_2) \in \hat{\mathbb{R}}^2 : |\xi_1| \geqslant \frac{1}{8}, \left|\frac{\xi_2}{\xi_1}\right| \leqslant 1\}$. That is, the functions $\{\hat{\psi}^{(0)}(\xi A_0^{-j} B_0^{-\ell})\}$ form a tiling of $\mathcal{D}_0$. This is illustrated in Fig. 1a.

This property, together with the fact that $\hat{\psi}^{(0)}$ is supported inside $\left[-\frac{1}{2}, \frac{1}{2}\right]^2$, implies that the collection:

$$\left\{\psi_{j,\ell,k}^{(0)}(x) = 2^{\frac{3j}{2}} \psi^{(0)}\left(B_0^\ell A_0^j x - k\right) : j \geqslant 0, \ -2^j \leqslant \ell \leqslant 2^j - 1, \ k \in \mathbb{Z}^2\right\}, \tag{2.6}$$

is a Parseval frame for $L^2(\mathcal{D}_0)^\vee = \{f \in L^2(\mathbb{R}^2) : \text{supp } \hat{f} \subset \mathcal{D}_0\}$. Details about this can be found in [22].

Similarly we can construct a Parseval frame for $L^2(\mathcal{D}_1)^\vee$, where $\mathcal{D}_1$ is the vertical cone $\mathcal{D}_1 = \{(\xi_1, \xi_2) \in \hat{\mathbb{R}}^2 : |\xi_2| \geqslant \frac{1}{8}, \left|\frac{\xi_1}{\xi_2}\right| \leqslant 1\}$. Let

$$A_1 = \begin{pmatrix} 2 & 0 \\ 0 & 4 \end{pmatrix}, \qquad B_1 = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix},$$

and $\psi^{(1)}$ be given by

$$\hat{\psi}^{(1)}(\xi) = \hat{\psi}^{(1)}(\xi_1, \xi_2) = \hat{\psi}_1(\xi_2)\hat{\psi}_2\left(\frac{\xi_1}{\xi_2}\right),$$

where $\hat{\psi}_1$, $\hat{\psi}_2$ are defined as above. Then the collection

$$\left\{\psi_{j,\ell,k}^{(1)}(x) = 2^{\frac{3j}{2}} \psi^{(1)}\left(B_1^\ell A_1^j x - k\right) : j \geqslant 0, \ -2^j \leqslant \ell \leqslant 2^j - 1, \ k \in \mathbb{Z}^2\right\} \tag{2.7}$$

is a Parseval frame for $L^2(\mathcal{D}_1)^\vee$. Finally, let $\hat{\varphi} \in C_0^\infty(\mathbb{R}^2)$ be chosen to satisfy

$$G(\xi) = \left|\hat{\varphi}(\xi)\right|^2 + \sum_{j \geqslant 0} \sum_{\ell=-2^j}^{2^j-1} \left|\hat{\psi}^{(0)}\left(\xi A_0^{-j} B_0^{-\ell}\right)\right|^2 \chi_{\mathcal{D}_0}(\xi) + \sum_{j \geqslant 0} \sum_{\ell=-2^j}^{2^j-1} \left|\hat{\psi}^{(1)}\left(\xi A_1^{-j} B_1^{-\ell}\right)\right|^2 \chi_{\mathcal{D}_1}(\xi) = 1$$

for $\xi \in \hat{\mathbb{R}}^2$,

where $\chi_{\mathcal{D}}$ denotes the indicator function of the set $\mathcal{D}$. This implies that $\operatorname{supp}\hat{\varphi} \subset \left[-\frac{1}{8}, \frac{1}{8}\right]^2$, with $|\hat{\varphi}(\xi)| = 1$ for $\xi \in \left[-\frac{1}{16}, \frac{1}{16}\right]^2$, and the set $\{\varphi(x-k): k \in \mathbb{Z}^2\}$ is a Parseval frame for $L^2\left(\left[-\frac{1}{16}, \frac{1}{16}\right]^2\right)^{\vee}$. Observe that, by the properties of $\psi^{(d)}$, $d = 0, 1$, it follows that the function $G(\xi) = G(\xi_1, \xi_2)$ is continuous and regular along the lines $\xi_2/\xi_1 = \pm 1$ (as well as for any other $\xi \in \hat{\mathbb{R}}^2$).

Thus, we have the following:

Theorem 2.1. *Let* $\varphi_k(x) = \varphi(x-k)$ *and* $\psi_{j,\ell,k}^{(d)}(x) = 2^{\frac{3j}{2}}\psi^{(d)}(B_d^{\ell}A_d^j x - k)$, *where* φ, ψ *are given as above. Then the collection of* shearlets:

$$\left\{\varphi_k: k \in \mathbb{Z}^2\right\} \cup \left\{\psi_{j,\ell,k}^{(d)}(x): j \geqslant 0, \ -2^j + 1 \leqslant \ell \leqslant 2^j - 2, \ k \in \mathbb{Z}^2, \ d = 0, 1\right\}$$

$$\cup \left\{\tilde{\psi}_{j,\ell,k}^{(d)}(x): j \geqslant 0, \ \ell = -2^j, 2^j - 1, \ k \in \mathbb{Z}^2, \ d = 0, 1\right\},$$

where $\hat{\tilde{\psi}}_{j,\ell,k}^{(d)}(\xi) = \hat{\psi}_{j,\ell,k}^{(d)}(\xi)\chi_{\mathcal{D}_d}(\xi)$, *is a Parseval frame for* $L^2(\mathbb{R}^2)$.

As shown above, the "corner" elements $\tilde{\psi}_{j,\ell,k}^{(d)}(x)$, $\ell = -2^j, 2^j - 1$, are simply obtained by truncation on the cones $\chi_{\mathcal{D}_d}$ in the frequency domain. As mentioned above, the corner elements in the horizontal cone $\mathcal{D}_0$ match nicely with those in the vertical cone $\mathcal{D}_1$.

For $d = 0, 1$, the *shearlet transform* is mapping $f \in L^2(\mathbb{R}^2)$ into the elements $\langle f, \psi_{j,\ell,k}^{(d)}\rangle$, where $j \geqslant 0$, $-2^j \leqslant \ell \leqslant 2^j - 1$, $k \in \mathbb{Z}^2$.

Let us summarize the mathematical properties of shearlets:

- Shearlets are *well localized*. In fact, they are compactly supported in the frequency domain and have fast decay in the spatial domain.
- Shearlets satisfy *parabolic scaling*. Each element $\hat{\psi}_{j,\ell,k}$ is supported on a pair of trapezoids, each one contained in a box of size approximately $2^j \times 2^{2j}$ (see Fig. 1b). Because the shearlets are well localized, in the spatial domain each $\psi_{j,\ell,k}$ is essentially supported on a box of size $2^{-j} \times 2^{-2j}$. Their supports become increasingly thin as $j \to \infty$.
- Shearlets exhibit *highly directional sensitivity*. The elements $\hat{\psi}_{j,\ell,k}$ are oriented along lines with slope given by $-\ell 2^{-j}$. As a consequence, the corresponding elements $\psi_{j,\ell,k}$ are oriented along lines with slope $\ell 2^{-j}$. The number of orientations doubles at each finer scale.
- Shearlets are *spatially localized*. For any fixed scale and orientation, the shearlets are obtained by translations on the lattice $\mathbb{Z}^2$.
- Shearlets are *optimally sparse*. The following is proved in [19, Theorem 1.1]):

Theorem. *Let* f *be* C^2 *away from piecewise* C^2 *curves, and* f_N^S *be the approximation to* f *obtained using the* N *largest coefficients in the shearlet expansion. Then we have*

$$\|f - f_N^S\|_2^2 \leqslant CN^{-2}(\log N)^3.$$

Thus the shearlets form a tight frame of well-localized waveforms, at various scales and directions, and are optimally sparse in representing images with edges. Only the curvelets of Candès and Donoho are known to satisfy similar sparsity properties.[2] With respect to the curvelets, however, our construction has some fundamental differences. Indeed, the shearlets are generated from the action of a family of operators on a single function, while this is not true for the curvelets (they are not of the form (2.2)). In particular, unlike the shearlets, the curvelets are not associated with a *fixed* translation lattice. Concerning the directional sensitivity, the number of orientations in our construction doubles at each scale, while in the curvelet case it doubles at each *other* scale. This is consistent with the fact that our

---

[2] Also the contourlets are claimed in [12] to satisfy the same sparsity property. The argument used in [12] assumes that there exist smooth compactly supported functions approximating a frequency partition similar to Fig. 1. However, the existence of functions with such properties is an open and non-trivial problem.

dilations factors in the dilation matrix A are 4 and 2 rather than 2 and $\sqrt{2}$, as in the case of curvelets. In addition, the shearlets are defined on the Cartesian domain and the various directions are obtained from the action of shearing transformations. By contrast, the curvelets are constructed in the polar domain and the orientations are obtained by applying rotations. Finally, thanks to their mathematical structure, the shearlets are associated to a multiresolution analysis (see [27,30]).

Also the *discrete* construction of the contourlets introduced by Do and Vetterli [12] has the intent to provide a partition of the frequency plane very similar to the one represented in Fig. 1. In this sense, the theory of shearlets can be seen as a theoretical justification for the contourlets. Observe, however, that the shearlets are band-limited functions, while the contourlets are a discrete-time construction implemented using filter banks. Indeed, the framework of composite wavelets from which the shearlets are derived allows one to consider directional multiscale representations with compact support [26]. It is an open problem whether one can construct a directional multiscale Parseval frame of functions that are both compactly supported and smooth.

3. The discrete shearlet transform

It will be convenient to describe the collection of shearlets presented above in a way which is more suitable to derive its numerical implementation. For $\xi = (\xi_1, \xi_2) \in \hat{\mathbb{R}}^2$, $j \geqslant 0$, and $\ell = -2^j, \ldots, 2^j - 1$, let

$$
W_{j,\ell}^{(0)}(\xi) = \begin{cases} \hat{\psi}_2(2^j \frac{\xi_2}{\xi_1} - \ell)\chi_{\mathcal{D}_0}(\xi) + \hat{\psi}_2(2^j \frac{\xi_1}{\xi_2} - \ell + 1)\chi_{\mathcal{D}_1}(\xi) & \text{if } \ell = -2^j, \\ \hat{\psi}_2(2^j \frac{\xi_2}{\xi_1} - \ell)\chi_{\mathcal{D}_0}(\xi) + \hat{\psi}_2(2^j \frac{\xi_1}{\xi_2} - \ell - 1)\chi_{\mathcal{D}_1}(\xi) & \text{if } \ell = 2^j - 1, \\ \hat{\psi}_2(2^j \frac{\xi_2}{\xi_1} - \ell) & \text{otherwise}, \end{cases}
$$

and

$$
W_{j,\ell}^{(1)}(\xi) = \begin{cases} \hat{\psi}_2(2^j \frac{\xi_2}{\xi_1} - \ell + 1)\chi_{\mathcal{D}_0}(\xi) + \hat{\psi}_2(2^j \frac{\xi_1}{\xi_2} - \ell)\chi_{\mathcal{D}_1}(\xi) & \text{if } \ell = -2^j, \\ \hat{\psi}_2(2^j \frac{\xi_2}{\xi_1} - \ell - 1)\chi_{\mathcal{D}_0}(\xi) + \hat{\psi}_2(2^j \frac{\xi_1}{\xi_2} - \ell)\chi_{\mathcal{D}_1}(\xi) & \text{if } \ell = 2^j - 1, \\ \hat{\psi}_2(2^j \frac{\xi_1}{\xi_2} - \ell) & \text{otherwise}, \end{cases}
$$

where $\psi_2, \mathcal{D}_0, \mathcal{D}_1$ are defined in Section 2. For $1 - 2^j \leqslant \ell \leqslant 2^j - 2$, each term $W_{j,\ell}^{(d)}(\xi)$ is a window function localized on a pair of trapezoids, as illustrated in Fig. 1a. When $\ell = -2^j$ or $\ell = 2^j - 1$, at the junction of the horizontal cone $\mathcal{D}_0$ and the vertical cone $\mathcal{D}_1$, $W_{j,\ell}^{(d)}(\xi)$ is the superposition of two such functions.

Using this notation, for $j \geqslant 0$, $-2^j \leqslant \ell \leqslant 2^j - 1$, $k \in \mathbb{Z}^2$, $d = 0, 1$, we can write the Fourier transform of the shearlets in the compact form

$$
\hat{\psi}_{j,\ell,k}^{(d)}(\xi) = 2^{\frac{3j}{2}} V(2^{-2j}\xi) W_{j,\ell}^{(d)}(\xi) e^{-2\pi i \xi A_d^{-j} B_d^{-\ell} k},
$$

where $V(\xi_1, \xi_2) = \hat{\psi}_1(\xi_1)\chi_{\mathcal{D}_0}(\xi_1, \xi_2) + \hat{\psi}_1(\xi_2)\chi_{\mathcal{D}_1}(\xi_1, \xi_2)$. The *shearlet transform* of $f \in L^2(\mathbb{R}^2)$ can be computed by

$$
\langle f, \psi_{j,\ell,k}^{(d)} \rangle = 2^{\frac{3j}{2}} \int\limits_{\mathbb{R}^2} \hat{f}(\xi) \overline{V(2^{-2j}\xi) W_{j,\ell}^{(d)}(\xi)} e^{2\pi i \xi A_d^{-j} B_d^{-\ell} k} \, d\xi. \tag{3.8}
$$

Indeed, one can easily verify that

$$
\sum_{d=0}^{1} \sum_{\ell=-2^j}^{2^j - 1} \left| W_{j,\ell}^{(d)}(\xi_1, \xi_2) \right|^2 = 1,
$$

and from this it follows that

$$
\left| \hat{\varphi}(\xi_1, \xi_2) \right|^2 + \sum_{d=0}^{1} \sum_{j \geqslant 0} \sum_{\ell=-2^j}^{2^j - 1} \left| V(2^{2j}\xi_1, 2^{2j}\xi_2) \right| \left| W_{j,\ell}^{(d)}(\xi_1, \xi_2) \right|^2 = 1 \quad \text{for } (\xi_1, \xi_2) \in \hat{\mathbb{R}}^2.
$$

3.1. A frequency-domain implementation

We will now derive an algorithmic procedure for computing (3.8) in frequency domain which is faithful to the mathematical transformation described above.

An $N \times N$ image consists of a finite sequence of values, $\{x[n_1, n_2]\}_{n_1,n_2=0}^{N-1,N-1}$ where $N \in \mathbb{N}$. Identifying the domain with the finite group $\mathbb{Z}_N^2$, the inner product of images $x, y : \mathbb{Z}_N^2 \to \mathbb{C}$ is defined as

$$\langle x, y \rangle = \sum_{u=0}^{N-1} \sum_{v=0}^{N-1} x(u, v) \overline{y(u, v)}.$$

Thus the discrete analog of $L^2(\mathbb{R}^2)$ is $\ell^2(\mathbb{Z}_N^2)$.

Given an image $f \in \ell^2(\mathbb{Z}_N^2)$, let $\hat{f}[k_1, k_2]$ denote its 2D discrete Fourier transform (DFT):

$$\hat{f}[k_1, k_2] = \frac{1}{N} \sum_{n_1,n_2=0}^{N-1} f[n_1, n_2] e^{-2\pi i (\frac{n_1}{N} k_1 + \frac{n_1}{N} k_2)}, \quad -\frac{N}{2} \leqslant k_1, k_2 < \frac{N}{2}.$$

Here and in the following we adopt the convention that brackets $[\cdot, \cdot]$ denote arrays of indices, and parentheses $(\cdot, \cdot)$ denote function evaluations. We shall interpret the numbers $\hat{f}[k_1, k_2]$ as samples $\hat{f}[k_1, k_2] = \hat{f}(k_1, k_2)$ from the trigonometric polynomial

$$\hat{f}(\xi_1, \xi_2) = \sum_{n_1,n_2=0}^{N-1} f[n_1, n_2] e^{-2\pi i (\frac{n_1}{N} \xi_1 + \frac{n_1}{N} \xi_2)}.$$

First, to compute

$$\hat{f}(\xi_1, \xi_2) \overline{V(2^{-2j} \xi_1, 2^{-2j} \xi_2)} \tag{3.9}$$

in the discrete domain, at the resolution level j, we apply the Laplacian pyramid algorithm [4], which is implemented in the time-domain. This will accomplish the multiscale partition illustrated in Fig. 1, by decomposing $f_a^{j-1}[n_1, n_2]$, $0 \leqslant n_1, n_2 < N_{j-1}$, into a low-pass filtered image $f_a^j[n_1, n_2]$, a quarter of the size of $f_a^{j-1}[n_1, n_2]$, and a high-pass filtered image $f_d^j[n_1, n_2]$. Observe that the matrix $f_a^j[n_1, n_2]$ has size $N_j \times N_j$, where $N_j = 2^{-2j} N$, and $f_a^0[n_1, n_2] = f[n_1, n_2]$ has size $N \times N$. In particular, we have that

$$\hat{f}_d^j(\xi_1, \xi_2) = \hat{f}(\xi_1, \xi_2) \overline{V(2^{-2j} \xi_1, 2^{-2j} \xi_2)}$$

and thus, $f_d^j[n_1, n_2]$ are the discrete samples of a function $f_d^j(x_1, x_2)$, whose Fourier transform is $\hat{f}_d^j(\xi_1, \xi_2)$.

In order to obtain the directional localization illustrated in Fig. 1, we will compute the DFT on the pseudo-polar grid, and then apply a one-dimensional band-pass filter to the components of the signal with respect to this grid. More precisely, let us define the pseudo-polar coordinates $(u, v) \in \mathbb{R}^2$ as follows:

$$(u, v) = \left(\xi_1, \frac{\xi_2}{\xi_1} \right) \quad \text{if } (\xi_1, \xi_2) \in \mathcal{D}_0,$$

$$(u, v) = \left(\xi_2, \frac{\xi_1}{\xi_2} \right) \quad \text{if } (\xi_1, \xi_2) \in \mathcal{D}_1.$$

After performing this change of coordinates, we obtain $g_j(u, v) = \hat{f}_d^j(\xi_1, \xi_2)$, and, for $\ell = 1 - 2^j, \ldots, 2^j - 1$, we have:

$$\hat{f}(\xi_1, \xi_2) \overline{V(2^{-2j} \xi_1, 2^{-2j} \xi_2) W_{j\ell}^{(d)}(\xi_1, \xi_2)} = g_j(u, v) \overline{W(2^j v - \ell)}. \tag{3.10}$$

This expression shows that the different directional components are obtained by simply translating the window function W. The discrete samples $g_j[n_1, n_2] = g_j(n_1, n_2)$ are the values of the DFT of $f_d^j[n_1, n_2]$ on a pseudo-polar grid. That is, the samples in the frequency domain are taken not on a Cartesian grid, but along lines across the origin at various slopes. This has been recently referred to as the *pseudo-polar* grid. One may obtain the discrete frequency values of f_d^j on the pseudo-polar grid by direct extraction using the fast Fourier transform (FFT) with complexity $O(N^2 \log N)$ or by using the pseudo-polar DFT (PDFT).

Definition 3.1. For a given $N \times N$ signal $f[n_1, n_2]$, the *pseudo-polar DFT* is given by $Pf = [\hat{f}_1, \hat{f}_2]^T$, where $\hat{f}_1, \hat{f}_2$ are given by

$$\hat{f}_1[k_1, k_2] = \sum_{n_1=-N/2}^{N/2-1} \sum_{n_2=-N/2}^{N/2-1} f[n_1, n_2] e^{-in_1 \frac{\pi k_1}{N}} e^{-in_2 \frac{\pi k_1}{N} \frac{2k_2}{N}},$$

for $-N \leqslant k_1 < N$, $-N/2 \leqslant k_2 < N/2$, and

$$\hat{f}_2[k_1, k_2] = \sum_{n_1=-N/2}^{N/2-1} \sum_{n_2=-N/2}^{N/2-1} f[n_1, n_2] e^{-in_2 \frac{\pi k_2}{N}} e^{in_1 \frac{\pi k_2}{N} \frac{2k_1}{N}},$$

for $-N/2 \leqslant k_1 < N/2$, $-N \leqslant k_2 < N$.

It is known that the operator P can be preconditioned so that it provides a nearly tight frame (when using a direct extraction routine the operator is a tight frame [10]). Furthermore, both the forward and inverse PDFT can be implemented with complexity $O(N^2 \log N)$ using the pseudo-polar FFT (see [2], where this is referred to as the *fast slant stack algorithm*).

Now let $\{w_{j,\ell}[n]: n \in \mathbb{Z}\}$ be the sequence whose discrete Fourier transform gives the discrete samples of the window function $\overline{W(2^j k - \ell)}$, that is, $\hat{w}_{j,\ell}[k] = \overline{W(2^j k - \ell)}$. Then, for fixed $n_1 \in \mathbb{Z}$, we have

$$\mathcal{F}_1\big(\mathcal{F}_1^{-1}\big(g_j[n_1, n_2]\big) * w_{j\ell}[n_2]\big) = g_j[n_1, n_2]\mathcal{F}_1\big(w_{j\ell}[n_2]\big), \tag{3.11}$$

where $*$ denotes the one-dimensional convolution along the n_2 axis. Here, $\mathcal{F}_1$ is the one-dimensional discrete Fourier transform defined by

$$\mathcal{F}_1(q)[k_1] = \frac{1}{\sqrt{N}} \sum_{n_1=-N/2}^{N/2-1} q[n_1] e^{\frac{-2\pi i k_1 n_1}{N}}$$

for a given 1-D signal q with length N. Thus (3.11) gives the algorithmic implementation for computing the discrete samples of $g_j(u, v)\overline{W(2^j v - \ell)}$.

Using the notation we have introduced, the shearlet coefficients $\langle f, \psi_{j,\ell,k}^{(0)}\rangle$, given by (3.8), are now simply

$$\int \int 2^{-\frac{3}{2}j} g_j(u, v) \overline{W(2^j v - \ell)} \exp\left(2\pi i \left(\frac{n_1 + \ell n_2}{4^j}\xi_1 + \frac{n_2}{2^j}\xi_2\right)\right) d\xi_1 d\xi_2. \tag{3.12}$$

Thus, to compute (3.12) in the discrete domain, it suffices to compute the inverse PDFT or directly re-assemble the Cartesian sampled values and apply the inverse two-dimensional FFT. For $d = 1$, the shearlet coefficients $\langle f, \psi_{j,\ell,k}^{(1)}\rangle$ are computed in similar way.

Observe that, in this implementation, we have a large flexibility in the choice of the frequency window function W. As we mentioned in Section 2 there are plenty of choices in the construction of the function ψ_2 (recall that W is defined in terms of the function ψ_2). In Section 3.3 we will implement the windowing in the time-domain. This way we will be able to create the windowing using wavelet filters by combining the decomposition and synthesis filters appropriately.

Let us summarize the procedure described above at fixed resolution level j. This is illustrated by the scheme of Fig. 2. Suppose $f \in \ell^2(\mathbb{Z}_N^2)$.

(1) Apply the Laplacian pyramid scheme to decompose f_a^{j-1} into a low-pass image f_a^j and a high-pass image f_d^j. For $f_a^{j-1} \in \ell^2(\mathbb{Z}_{N_{j-1}}^2)$, the matrix $f_a^j \in \ell^2(\mathbb{Z}_{N_j}^2)$, where $N_j = N_{j-1}/4$ and $f_d^j \in \ell^2(\mathbb{Z}_{N_{j-1}}^2)$.

(2) Compute $\hat{f}_d^j$ on a pseudo-polar grid. This gives the matrix Pf_d^j.

(3) Apply a band-pass filtering to the matrix Pf_d^j (this performs (3.11)).

(4) Directly re-assemble the Cartesian sampled values and apply the inverse two-dimensional FFT or use the inverse PDFT from the previous step.
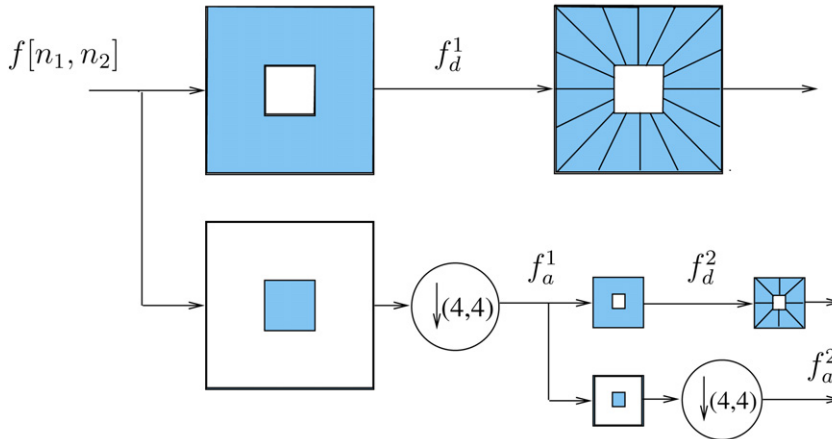
Fig. 2. The figure illustrates the succession of Laplacian pyramid and directional filtering.

The algorithm runs in $O(N^2 \log N)$ operations.

Figure 3 illustrates the two level shearlet decomposition of the Peppers image. The first level decomposition generates 4 subbands, and the second level decomposition generates 8 subbands, corresponding to the different directional bands illustrated by the scheme in Fig. 2.

Figure 4 displays examples of the basis functions for the frequency-domain based shearlet transform. The first level decomposition is separated into 16 directional subbands and the second level decomposition is separated into 8 directional subbands.

3.2. Correlation with the theory

The above frequency-based implementation yields the spatial-frequency tiling determined by the shearlet transform. Recall that each element $\hat{\psi}_{j,\ell,k}$ is supported on a pair of trapezoids, each one contained in a box of size approximately $2^j \times 2^{2j}$. Thus, the non-linear approximation error rate is expected to be what the theory predicts. Note that, in this implementation, the downsampling is only applied in the vertical and horizontal directions with no anisotropic subsampling. Thus the decomposition is highly redundant. For example, given an image of size $N \times N$, a three-level decomposition would contain $2^j N^2 + 2^{j-1}(N/4)^2 + (N/16)^2$ coefficients when 2^j directional subbands are chosen at the first decomposition level. The incorporation of anisotropic subsampling is a non-trivial matter that will be investigated in a follow-up paper.

We now demonstrate that the approximation properties predicted by the theory of shearlets are very closely correlated to the corresponding properties of its discrete implementation.

Figure 5b shows the non-linear approximation error $\|f - f_M\|/\|f\|$ where f_M is the partial reconstruction of f using the M-largest coefficients in the shearlet (or wavelet) representation. In order to compensate for the redundancy of the shearlet transform and display a fair comparison, we multiply the number of wavelet coefficients by the redundancy factor of the shearlet transform so that the number of shearlet and wavelet coefficients is identical.

Our second test image shown in Fig. 6a is singular along smooth circles and is otherwise smooth. Our theory (1.1) tells us that in this case we have

$$\|f - f_M\| \leqslant C_p M^{-1+p}$$

for any $p > 0$ so that the decay rate of the non-linear approximation curve is at least arbitrarily close to 1. Our numerical experiment (see Fig. 6b) shows that the decay rate of the non-linear approximation curve for the shearlet transform is close to 1 for our test image shown in Fig. 5a. In this numerical experiment, we compare the non-linear approximation curve for our shearlet representation and the numerically estimated curve of the form $CM^{-\alpha}$. For this estimated curve, we obtained $\alpha \simeq 0.9634$ which is close to 1.

Although there is a redundancy in the number of retained coefficients, the asymptotic decay rate demonstrated above indicates that this discrete implementation should perform well as a denoising routine. An analogous situation
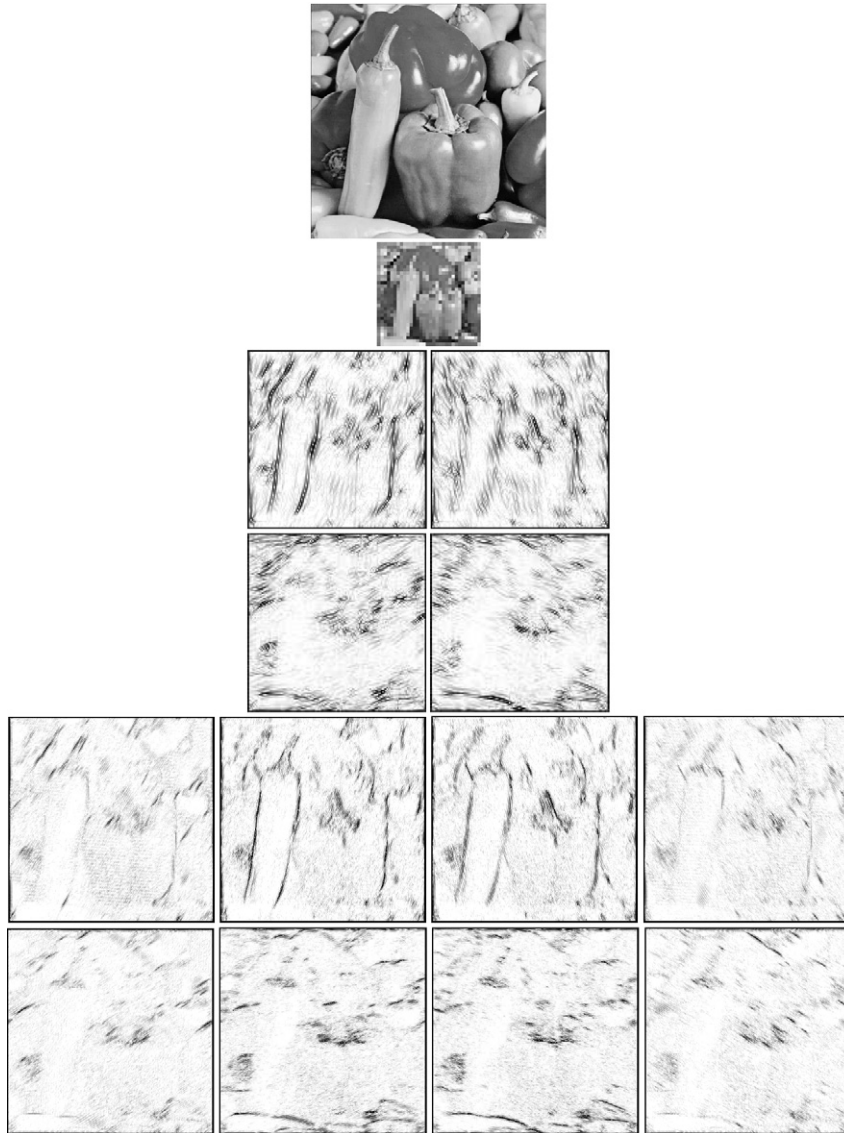
Fig. 3. An illustration of the shearlet transform. The top image is the original Peppers image. The image below the top image contains the approximate shearlet coefficients. Images of the detail shearlet coefficients are shown below this with an inverted grayscale for better presentation.

occurs for the wavelet transform and its implementation. The success of the wavelet transform for denoising is based on its non-linear approximation error rate and yet their most successful implementations for estimation purposes are typically done by using the highly redundant non-subsampled version (see, for example, [8,28]).

3.3. A time-domain implementation

In order to improve the algorithm performance for applications such as denoising, we need to implement a local variant of the shearlet transform. This will reduce the Gibbs type ringing present when filters of large support sizes are used. Note that the concept of localizing the transform is not new. For example, a localization has been applied to the ridgelet transform in order to implement the discrete curvelet transform [35].

In order to obtain a local variant similar to the one used for ridgelets, we would need to apply the shearlet transforms to small sized image blocks (e.g., blocks of sizes 8 by 8, or 16 by 16). In order to avoid blocking artifacts, we would need to introduce an overcomplete decomposition of the image and then synthesize by a lapped window scheme such
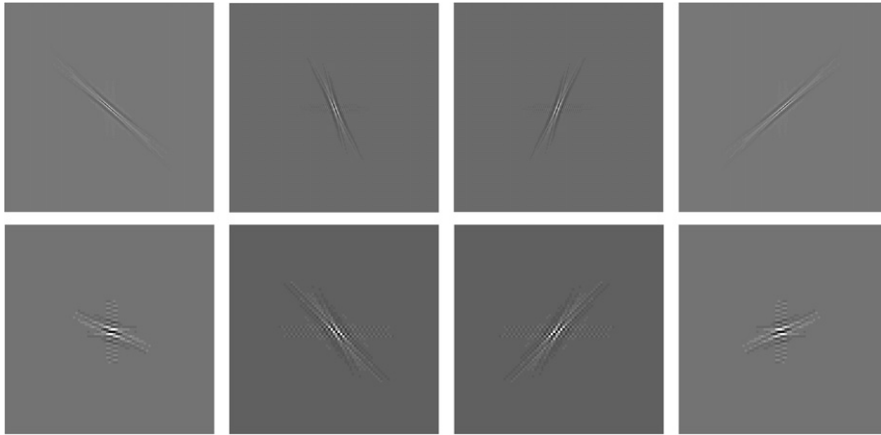
Fig. 4. Examples of basis functions of the frequency-domain implemented shearlet transform. The top row corresponds to the basis functions of the first decomposition level. The bottom row corresponds to the basis functions of the second decomposition level.
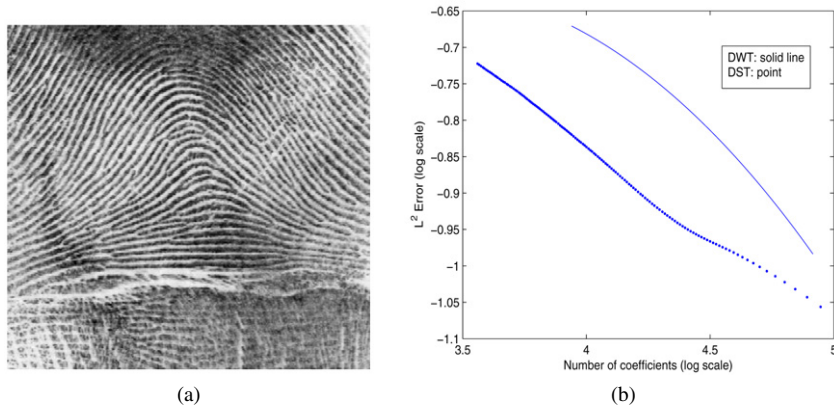

Fig. 5. Non-linear approximation curve of the shearlet and wavelet representations. (a) Original image. (b) Partial reconstruction error $\|f - f_M\|/\|f\|$.
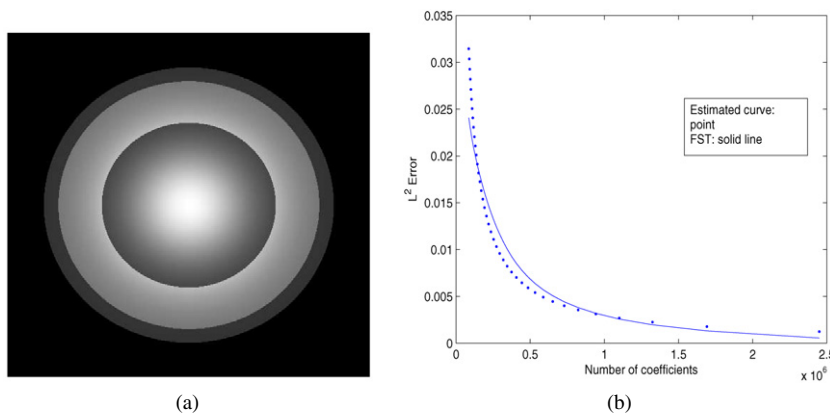

Fig. 6. (a) Original image. (b) Partial reconstruction error $\|f - f_M\|/\|f\|$ and the numerically estimated curve.

as in [35]. Indeed, the added redundancy in the number of shearlet coefficients makes this a cumbersome approach. But a much simpler, faster, and time-domain solution is possible.

Recall that in the implementation there is a large flexibility in the choice of windowing to be applied. Consider a frequency-based window function $\tilde{W}$ such that $\sum_{\ell=-2^j}^{2^j-1} \tilde{W}[2^j n_2 - \ell] = 1$. Denote by φ_P the mapping function from the Cartesian grid to the pseudo-polar grid. The shearlet coefficients in the discrete Fourier domain were earlier cal-
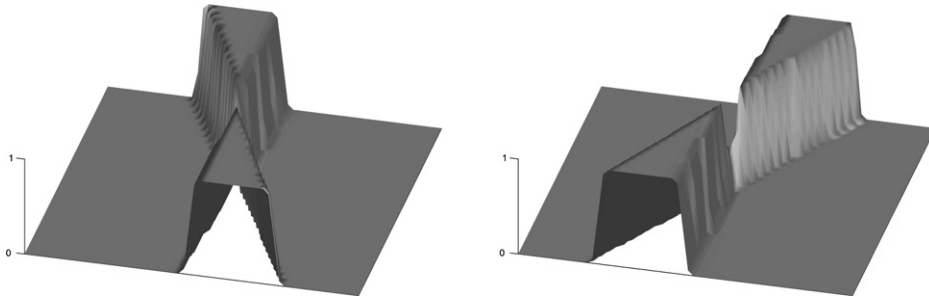
Fig. 7. Examples of the shearing filters $\hat{w}_{j,\ell}^s$ constructed using a Meyer wavelet as the window function W.

culated as $\varphi_P^{-1}(g_j[n_1, n_2]\tilde{W}[2^j n_2 - \ell])$, where g_j represents $\hat{f}_d^j$ in the pseudo-polar domain. We suggest to calculate the shearlet coefficients in the frequency domain as

$$\varphi_P^{-1}\big(g_j[n_1, n_2]\big)\varphi_P^{-1}\big(\hat{\delta}_P[n_1, n_2]\tilde{W}[2^j n_2 - \ell]\big),$$

where $\hat{\delta}_P$ represents the discrete Fourier transform of the delta function in the pseudo-polar grid. This is possible because the map φ_P can be described as a *selection matrix S* with the property that its elements $s_{i,j}$ satisfy the property $s_{i,j}^2 = s_{i,j}$ (see [10] for more details). In this way, we can calculate the shearlet coefficients in the discrete Fourier domain as $\hat{f}_d^j[n_1, n_2]\hat{w}_{j,\ell}^s[n_1, n_2]$ where

$$\hat{w}_{j,\ell}^s[n_1, n_2] = \varphi_P^{-1}\big(\hat{\delta}_P[n_1, n_2]\tilde{W}[2^j n_2 - \ell]\big).$$

The subtle but very important point here is that the new form of the filters are not found by a simple change of variables. They are found by applying the specific discrete re-sampling transformation converting from the pseudo-polar to Cartesian coordinate system. This discrete transformation requires a re-sampling, where many points in the pseudo-polar coordinate systems may be mapped into a single point in the Cartesian system.

An illustration of such filters $\hat{w}_{j,\ell}^s$ found by using a Meyer window are shown in Fig. 7.

As a result of this conversion we now have filters $\hat{w}_{j,\ell}^s$ such that

$$\sum_{\ell=-2^j}^{2^j-1} \hat{w}_{j,\ell}^s(\xi_1, \xi_2) = 1.$$

Because this construction is independent of the image f, we can now construct shearing filters for any size coordinate system. By taking the inverse discrete Fourier transform, we thus have the following theorem.

Theorem 3.1. *Let w_ℓ^s denote the shearing filter $w_{0,\ell}^s$ with support size $L \times L$. Given any function $f \in \ell^2(\mathbb{Z}_N^2)$,*

$$\sum_{\ell=-2^j}^{2^j-1} f * w_\ell^s = f.$$

Although these filters are not compactly supported in the traditional sense, they can be implemented with a matrix representation that is smaller in size than the given image.

These observations show that we can perform the shearing filtering "directly" in the time-domain using a convolution. In our implementation, we will restrict the convolution to be of the same size as the given image. In addition, the small support sizes of the filters reduce the Gibbs-type ringing phenomenon and improve the computational efficiency of the algorithm. In fact, the small sized filters allows us to use a fast overlap-add method to compute the convolutions [32]. The gain in speed for the directional filtering combined with the performance of the Laplacian pyramid algorithm used in our routine yields an overall performance of $O(N^2 \log N)$ operations.

Another benefit of this implementation is that we can apply a non-subsampled Laplacian pyramid decomposition which has been shown to be very effective in denoising applications [11]. Although this is a highly redundant decomposition (e.g., the number of retained coefficients for a three-level decomposition would be $(2^j + 2^{j-1} + 1)N^2$ when
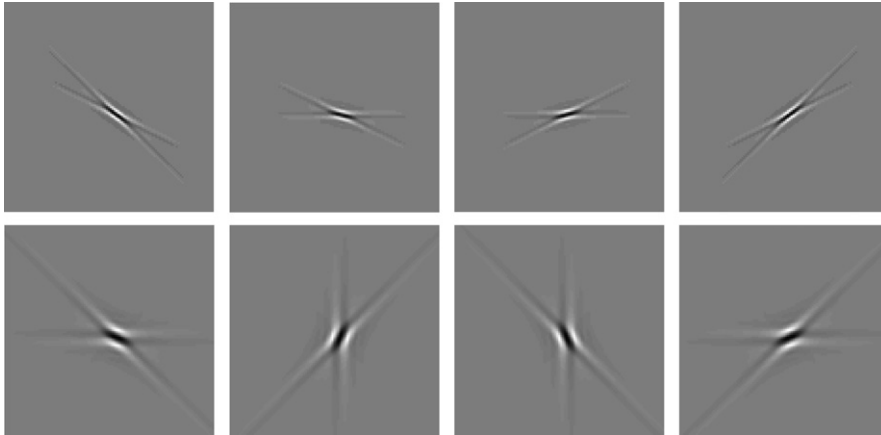
Fig. 8. Examples of basis functions of the time-domain implemented shearlet transform. The top row corresponds to the basis functions of the second decomposition level. The bottom row corresponds to the basis functions of the third decomposition level.

2^j directional subbands are chosen at the first decomposition level when applied to an $N \times N$ image), this version will be shown to be highly effective for the purpose of denoising.

On the other hand, it is useful to observe that the frequency-domain implementation discussed in the previous section allows for a much broader class of wavelet filtering (windowing) to be implemented. This can be useful for other types of applications.

Displays of various basis functions for the time-domain implemented shearlet transform are shown in Fig. 8. The second level decomposition was divided into 16 directional subbands and the third level decomposition was divided into 8 directional subbands.

3.3.1. Comparison with the contourlets

The time-domain shearlet transform that we described above has similarities with the contourlet transform [11,31, 33]. Recall that the contourlet transform consists of an application of the Laplacian pyramid followed by directional filtering. However the directional filtering is obtained using a different approach from the shearlets. Indeed, the directional filtering of the contourlet transform is achieved by introducing a directional filter bank that combines critically sampled fan filter banks and pre/post re-sampling operations.

An important advantage of the shearlet transform over the contourlet transform is that there are no restrictions on the number of directions for the shearing. That is, we could express the formulation of the windowing *W* with a non-dyadic spacing as well. This flexibility is not possible using a fan filter implementation. In addition, in the shearlet approach, there are no constraints on the size of the supports for the shearing, unlike the construction of the directional filter banks in [33]. Finally, we wish to point out that the inversion of this discrete shearlet transform only requires a summation of the shearing filters rather than inverting a directional filter bank. This results in an implementation that is most efficient computationally. In addition, this efficient inversion may have advantages for applications such as compression routines where the complexity of the decompression algorithm needs to be minimal.

3.4. Computational efficiency and accuracy

To give an indication of how computationally efficient the shearlet transform is, we have compared CPU times for computing the shear filtered coefficients (SFC) and its inversion processes (iSFC) to those of the non-subsampled directional filter bank (DFB) and its inversion process (iDFB) used as part of the non-subsampled contourlet transform.

Our test was based on using a laptop with a 1.73 GHz Centrino processor and 1 GB of RAM. The routines were tested in MATLAB with only one routine of the DFB codes compiled from C. The DFB codes were provided by the authors of the non-subsampled contourlet transform papers. The sizes of the shearing filters used were 16×16. We measured the following CPU times averaged over 10 iterations.

	Directions	Image size	CPU time
SFC	8	512	1.4641×10^0 s
iSFC	8	512	2.9687×10^{-2} s
SFC	16	512	2.9266×10^0 s
iSFC	16	512	1.0938×10^{-1} s
DFB	8	512	1.0850×10^2 s
iDFB	8	512	1.0867×10^2 s
DFB	16	512	2.7685×10^2 s
iDFB	16	512	2.7865×10^2 s

Notice that doubling the number of directions for the shearing processes only marginally increases the computational time, whereas it more than doubles the time for the directional filter bank process. Also, it can be seen that the time to invert the shearing is practically negligible for an image of size 512×512.

Below are the results for the frequency-domain shearing.

	Directions	Image size	CPU time
Freq-SFC	8	512	3.8898×10^1 s
Freq-iSFC	8	512	2.8125×10^{-1} s
Freq-SFC	16	512	1.2950×10^2 s
Freq-iSFC	16	512	5.8203×10^{-1} s

The average CPU time to decompose an image of size 512 via the LP algorithm used in the frequency-domain implementation for the shearlet transform was 3.2656×10^{-1}. The average CPU time to recompose via the LP algorithm was 3.2969×10^{-1}. For the non-subsampled LP algorithm used in the time-domain shearlet transform, the average CPU time was 4.7656×10^{-1}. The average CPU time to recompose was 4.8281×10^{-1}.

The relative error in the reconstruction of the frequency-domain implementation for an image of size 512 (Lena) was 3.8842×10^{-13}. For the non-subsampled time-domain shearlet transform, the relative errors were 7.8228×10^{-16} using a Meyer wavelet-based window and 7.8249×10^{-16} using a characteristic function based window. These results are acceptable and expected when implemented in a finite precision machine. The frequency-domain implementation only suffers from a slight performance degradation due to the limits on the discretization of the one-dimensional Meyer wavelet being used. The degradation is more visible when used on a signal of length 1024 than for a signal of length 32 or 64. Alternative discretizations of the Meyer wavelet could be used to mitigate this issue.

4. Numerical experiments

The highly directional sensitivity of the shearlet transform and its optimal approximation properties will lead to improvements in many image processing applications. To illustrate one of its potential uses, we have used the shearlet transform to remove noise from images. Specifically, suppose that for a given image f, we have

$$u = f + \epsilon, \tag{4.13}$$

where ϵ is Gaussian white noise with zero mean and standard deviation σ; that is, $\epsilon \in N(0, \sigma^2)$. We attempt to recover the image f from the noisy data u by computing an approximation $\tilde{f}$ of f obtained by applying a thresholding scheme in the subbands of the shearlet decomposition.

First, we demonstrate the performance in estimation by applying hard thresholding to the subbands of the shearlet decomposition using the frequency-based routine. The decomposition tested is the same as that shown in Fig. 3. The result is shown in Fig. 9. The performance measure used was the peak signal-to-noise ratio (PSNR) in decibels (dB) defined as

$$PSNR = 20 \log_{10} \frac{255N}{\|f - \tilde{f}\|_F},$$

where $\| \cdot \|_F$ is the Frobenius norm, the given image f is of size $N \times N$, and $\tilde{f}$ denotes the estimated image. Included in this experiment are the estimates found by applying hard thresholding to the discrete wavelet transform defined

Fig. 9. (a) The original cameraman image. (b) The noisy image (PSNR = 22.09 dB). (c) The result of wavelet denoising using the 7/9 fil-ters (PSNR = 26.18 dB). (d) The result of contourlet denoising (PSNR = 25.82 dB). (e) The result of the frequency-based shearlet denoising (PSNR = 27.21 dB). (f) The result of the time-domain shearlet denoising (PSNR = 28.01 dB).

in terms of the Daubechies–Antonini 7/9 filters and the contourlet transform using a decomposition compatible with the shearlet decomposition. We also include the performance when hard thresholding is applied to the time-domain based shearlet transform. The decomposition is the same as for the frequency-based transform but with the shearing

Fig. 10. Test images. From top left, clockwise: *Lena, Peppers, Elaine*, and *Goldhill*.

filters used to obtain the 8 and 4 directional subbands constructed using filters of size 16 and 32, respectively. This experiment suggests a better performance in using the time-domain shearlet transform for denoising and hence we provide a more complete set of comparisons using this implementation.

Taking note of the great performance of the non-subsampled contourlet transform for image denoising [11], we use a time-domain shearlet transform and choose the threshold parameters

$$\tau_{i,j} = \sigma^2_{\epsilon_{i,j}} / \sigma^2_{i,j,n} \tag{4.14}$$

as in [11] where $\sigma^2_{i,j,n}$ denotes the variance of the nth coefficient at the ith shearing direction subband in the jth scale, and $\sigma^2_{\epsilon_{i,j}}$ is the noise variance at scale j and shearing direction i. Various experiments indicate the shearlet coefficients can be modeled by generalized Gaussian distributions so that these thresholds should yield a risk close to the optimal Bayes risk, specifically within 5 percent of it. To estimate the signal variances in each subband locally, the neighboring coefficients contained in a square window and a maximum likelihood estimator are used. The variances $\sigma^2_{\epsilon_{i,j}}$ are estimated by using a Monte Carlo technique in which the variances are computed for several normalized noise images and then the estimates are averaged.

The particular form of the time-domain based shearlet transform we tested was to use the non-subsampled Laplacian pyramid transform with several different combinations of the shearing filters. This will be simply referred to as the non-subsampled shearlet transform (NSST). We use the abbreviation of $\mathrm{NSST}_1(L_1,L_2)$ and $\mathrm{NSST}_2(L_1,L_2)$ to indicate the type of windowing used and the support sizes of the shearing filters w^s_ℓ. In particular, we implemented the shearing on 4 of the 5 scales of the Laplacian pyramid transform decomposition. The shearing filters of sizes $L_1 \times L_1$, $L_1 \times L_1$, $L_2 \times L_2$, and $L_2 \times L_2$ from finer to coarser were used with the number of shearing directions chosen to be 16, 16, 8, and 8. Note that the only restriction on the construction of the shearing filters is that the maximum number

Table 1
Comparison of the perfomances of the shearlet denoising algorithm NSST$_1$ to other algorithms

	Noisy	BivShrink	DDWT	Curvelet	NSCT	NSST$_1$(16, 32)
Lena						
$\sigma = 10$	28.14 dB	34.36 dB	35.36 dB	33.71 dB	35.29 dB	35.38 dB
$\sigma = 15$	24.61 dB	32.48 dB	33.63 dB	32.52 dB	33.57 dB	33.71 dB
$\sigma = 20$	22.12 dB	31.16 dB	32.37 dB	31.54 dB	32.33 dB	32.47 dB
$\sigma = 25$	20.18 dB	30.16 dB	31.38 dB	30.66 dB	31.33 dB	31.46 dB
Peppers						
$\sigma = 10$	28.14 dB	33.51 dB	34.20 dB	32.81 dB	34.22 dB	34.35 dB
$\sigma = 15$	24.61 dB	31.98 dB	32.74 dB	31.72 dB	32.78 dB	32.97 dB
$\sigma = 20$	22.12 dB	30.80 dB	31.64 dB	30.84 dB	31.67 dB	31.90 dB
$\sigma = 25$	20.18 dB	29.87 dB	30.74 dB	30.01 dB	30.75 dB	30.99 dB
Goldhill						
$\sigma = 10$	28.14 dB	32.28 dB	32.86 dB	30.98 dB	32.87 dB	32.91 dB
$\sigma = 15$	24.61 dB	30.46 dB	31.17 dB	29.90 dB	31.14 dB	31.21 dB
$\sigma = 20$	22.12 dB	29.24 dB	30.00 dB	29.08 dB	29.97 dB	30.05 dB
$\sigma = 25$	20.18 dB	28.35 dB	29.13 dB	28.41 dB	29.09 dB	29.17 dB
Elaine						
$\sigma = 10$	28.14 dB	31.91 dB	32.83 dB	32.11 dB	32.86 dB	33.06 dB
$\sigma = 15$	24.61 dB	31.50 dB	31.79 dB	31.43 dB	31.84 dB	31.93 dB
$\sigma = 20$	22.12 dB	30.38 dB	31.09 dB	30.81 dB	31.15 dB	31.20 dB
$\sigma = 25$	20.18 dB	29.79 dB	30.51 dB	30.24 dB	30.55 dB	30.59 dB

of directional subbands is less than or equal to the size of the filter. NSST$_1$ refers to the case where the shearing was done by using a Meyer wavelet window and NSST$_2$ to the case where the shearing was done with a simple characteristic window function. For example, NSST$_1$(16, 32) indicates that a Meyer-based shearing filter of size 16 with 16 directions was applied to the first and second decomposition level and a Meyer-based shearing filter of size 32 with 8 directions was applied to the third and fourth decomposition level.

We tested the denoising schemes using the images shown in Fig. 10 for various standard deviation values of the noise. For a baseline comparison, we tested the performance of the standard discrete wavelet transform (DWT) and the stationary wavelet transform (SWT) both defined in terms of the Daubechies–Antonini 7/9 filters using hard thresholding. For brevity, the performance of these transforms using soft thresholding are not presented since they performed significantly less than the results obtained by hard thresholding. For more competitive comparisons, we tested the bivariate shrinkage algorithm (BivShrink), a thresholding technique based on taking into account the statistical dependencies among wavelet coefficients, using the discrete wavelet transform and using the dual-tree discrete wavelet transform (DDWT) [34]. We also compared the scheme against the curvelet based denoising scheme of [35] and the non-subsampled contourlet transform (NSCT) denoising scheme of [11] using 16, 16, 8, and 8 directions from finer to coarser scales.

The performance of the shearlet approach relative to other transforms is shown in Tables 1 and 2. It shows that the shearlet algorithm consistently outperforms all the algorithms mentioned above. NSST$_1$(16, 32) shows a fraction of a dB in improvement in terms of PSNR over the BivShrink and NSCT algorithms. The improvement over curvelets and wavelets is in many cases 1 dB or more. The improvement over close-ups of some of the best performing estimates are shown in Figs. 11 and 12 where it can be seen that these slight improvements are visually noticeable. The shearlet transform results exhibits less Gibbs-type residual artifacts than the other denoising methods. We attribute this to the small support sizes of the shearing filters.

5. Conclusion

We have developed both a frequency and time-domain based implementation of the discrete shearlet transform. These two different versions (although there is some commonality between them) were created for greater flexibility with future applications in mind. The frequency-based implementation gives much greater flexibility in the type of windowing that can be utilized and allows for the possibility of incorporating subsampling. This can be useful for

Table 2
Additional comparison of the perfomances of the shearlet denoising algorithms $NSST_1$ and $NSST_2$ to other algorithms

	DWT	SWT	$NSST_1(8, 16)$	$NSST_2(16, 32)$	$NSST_2(8, 16)$
Lena					
$\sigma = 10$	31.91 dB	33.73 dB	35.22 dB	35.23 dB	35.08 dB
$\sigma = 15$	30.10 dB	31.90 dB	33.56 dB	33.60 dB	33.47 dB
$\sigma = 20$	28.79 dB	30.55 dB	32.34 dB	32.39 dB	32.26 dB
$\sigma = 25$	27.79 dB	29.51 dB	31.35 dB	31.41 dB	31.28 dB
Peppers					
$\sigma = 10$	31.73 dB	33.30 dB	34.21 dB	34.24 dB	34.13 dB
$\sigma = 15$	29.96 dB	31.76 dB	32.83 dB	32.91 dB	32.77 dB
$\sigma = 20$	28.70 dB	30.57 dB	31.76 dB	31.86 dB	31.72 dB
$\sigma = 25$	27.70 dB	29.53 dB	30.87 dB	30.97 dB	30.83 dB
Goldhill					
$\sigma = 10$	29.67 dB	31.26 dB	32.76 dB	32.69 dB	32.56 dB
$\sigma = 15$	28.01 dB	29.53 dB	31.09 dB	31.08 dB	30.92 dB
$\sigma = 20$	26.95 dB	28.35 dB	29.94 dB	29.95 dB	29.81 dB
$\sigma = 25$	26.18 dB	27.48 dB	29.08 dB	29.10 dB	28.97 dB
Elaine					
$\sigma = 10$	30.72 dB	31.75 dB	32.67 dB	32.67 dB	32.56 dB
$\sigma = 15$	29.66 dB	30.75 dB	31.76 dB	31.79 dB	31.73 dB
$\sigma = 20$	28.78 dB	29.96 dB	31.11 dB	31.15 dB	31.09 dB
$\sigma = 25$	28.08 dB	29.30 dB	30.53 dB	30.58 dB	30.52 dB

Fig. 11. Close-up of images. From top left, clockwise: Noisy image (PSNR = 22.12 dB), BivShrink DDWT (PSNR = 31.09 dB), $NSST_1(16, 32)$ (PSNR = 31.20 dB), and NSCT (PSNR = 31.15 dB).
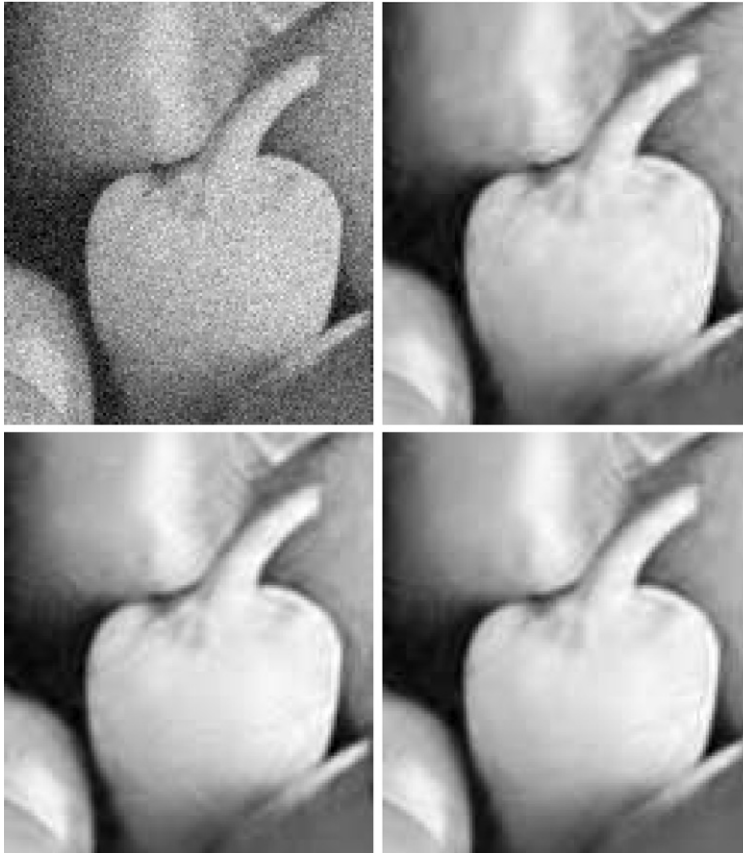
Fig. 12. Close-up of images. From top left, clockwise: Noisy image (PSNR = 22.12 dB), BivShrink DDWT (PSNR = 31.64 dB), NSST$_1$(16, 32) (PSNR = 31.90 dB), and NSCT (PSNR = 31.67 dB).

compression-type applications. For the time-domain based or finitely supported filtering implementation, the discrete shearlet transform becomes suitable for applications requiring translation invariance such as denoising and computational efficiency.

These discrete shearlet implementations are related to the discrete curvelet and contourlet transforms. All of these have a similar idealized frequency decomposition but differ in their implementation and construction. In fact, we noticed in various denoising experiments that the residual artifacts after reconstructions are very similar in nature. The features of each particular representation will have various advantages for specific applications. Take, for example, the use of the curvelet transform for image deconvolution to reduce computational complexity as demonstrated in [16].

In this paper, we have succeeded in demonstrating that the shearlet transform can be very competitive in performance for denoising images. The main advantages are that the shearing filters can have smaller support sizes than the directional filters used in the contourlet transform and can be implemented much more efficiently. We believe the small support sizes of the shearing filters may have been the reason for the slight improvement over the contourlet transform for the results tested as can be seen in the close up of the images shown. An additional appealing point to make in favor of the shearlets approach is that theoretically they transition very nicely from a continuous perspective to a discrete perspective. In addition, the proposed framework is suitable to many variations and generalizations.

In light of our developments in this work, other image and multidimensional data applications will benefit greatly with the use of the discrete shearlet transform. We intend to study some of these uses in future research endeavors.

Acknowledgments

Appendix A. Construction of ψ_1, ψ_2

In this appendix we show how to construct examples of functions ψ_1, ψ_2 satisfying the properties described in Section 2. Some ideas of these constructions are adapted from [19].

In order to construct ψ_1, let $h(t)$ be an even C_0^∞ function, with support in $\left(-\frac{1}{6}, \frac{1}{6}\right)$, satisfying $\int_{\mathbb{R}} h(t)\, \mathrm{d}t = \frac{\pi}{2}$, and define $\theta(\omega) = \int_{-\infty}^{\omega} h(t)\, \mathrm{d}t$. Then one can construct a smooth bell function as

$$b(\omega) = \begin{cases} \sin(\theta(|\omega| - \frac{1}{2})) & \text{if } \frac{1}{3} \leqslant |\omega| \leqslant \frac{2}{3}, \\ \sin\left(\frac{\pi}{2} - \theta\left(\frac{|\omega|}{2} - \frac{1}{2}\right)\right) & \text{if } \frac{2}{3} < |\omega| \leqslant \frac{4}{3}, \\ 0 & \text{otherwise.} \end{cases}$$

It follows from the assumptions we made (cf. [23, Section 1.4]) that

$$\sum_{j=-1}^{\infty} b^2\left(2^{-j}\omega\right) = 1 \quad \text{for } |\omega| \geqslant \frac{1}{3}.$$

Now letting $u^2(\omega) = b^2(2\omega) + b^2(\omega)$, it follows that

$$\sum_{j \geqslant 0}^{\infty} u^2\left(2^{-2j}\omega\right) = \sum_{j=-1}^{\infty} b^2\left(2^{-j}\omega\right) = 1 \quad \text{for } |\omega| \geqslant \frac{1}{3}.$$

Finally, let ψ_1 be defined by $\hat{\psi}_1(\omega) = u\left(\frac{8}{3}\omega\right)$. Then $\operatorname{supp} \hat{\psi}_1 \subset \left[-\frac{1}{2}, -\frac{1}{16}\right] \cup \left[\frac{1}{16}, \frac{1}{2}\right]$ and Eq. (2.4) is satisfied. This construction is illustrated in Fig. 13a.

For the construction of ψ_2, we start by considering a smooth bump function $f_1 \in C_0^\infty(-2, 2)$ such that $0 \leqslant f_1 \leqslant 1$ on $(-2, 2)$ and $f_1 = 1$ on $[-1, 1]$ (cf. [24, Section 1.4]). Next, let $f_2(t) = \sqrt{1 - e^{1/t}}$. Then (in the left-limit sense) $f_2(0) = 1$, $f_2^{(k)}(0) = 0$, for $k \geqslant 1$ and $0 < f_2 < 1$ on $(-1, 0)$. Define $f(t) = f_1(t-1) f_2(t-1)$, for $t \in [-1, 1]$. It is
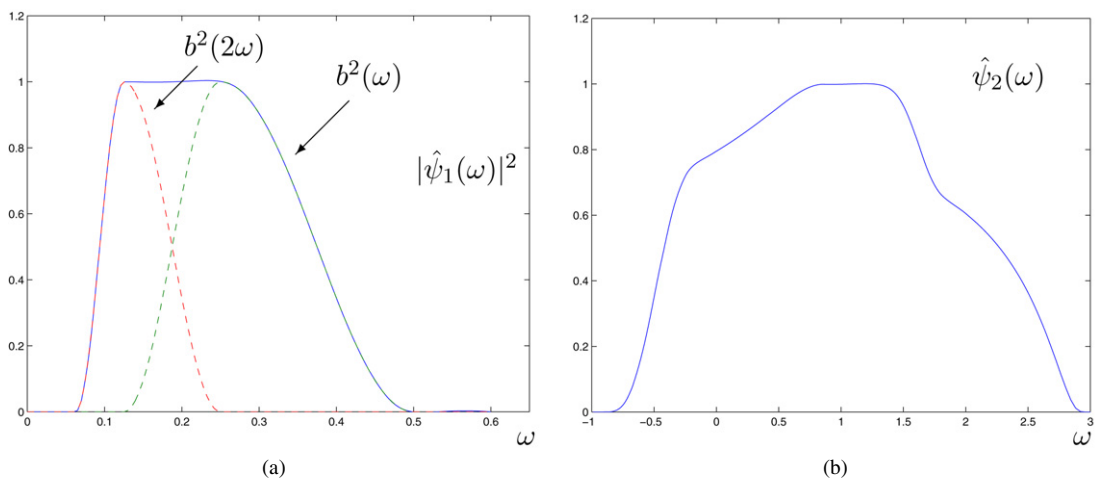

Fig. 13. (a) The function $|\hat{\psi}_1(\omega)|^2$ (solid line), for $\omega > 0$; the negative side is symmetrical. This function is obtained, after rescaling, from the sum of the window functions $b^2(\omega) + b^2(2\omega)$ (dashed lines). (b) The function $\hat{\psi}_2(\omega)$.

then easy to see that $f^{(k)}(-1) = 0$ for $k \geqslant 0$, $f(1) = 1$, and $f^{(k)}(1) = 0$ for $k \geqslant 1$. Let $g(t) = \sqrt{1 - f^2(t-2)}$. Since $g(t) = e^{\frac{1}{2(t-3)}}$, for $t \in (2, 3)$, it follows that $\lim_{t \to 3^-} g^{(k)}(t) = 0$, for $k \geqslant 0$. Finally, we define

$$\hat{\psi}_2(\omega) = \begin{cases} f(\omega) & \text{if } \omega \in [-1, 1), \\ g(\omega) & \text{if } \omega \in [1, 3], \\ 0 & \text{otherwise.} \end{cases}$$

Then $\hat{\psi}_2 \in C_0^\infty(\mathbb{R})$, with supp $\hat{\psi}_2 \subset [-1, 3]$, and

$$\hat{\psi}_2^2(\omega) + \hat{\psi}_2^2(\omega + 1) = 1, \quad \omega \in [-1, 1]. \tag{A.1}$$

From (A.1), it follows that, for any $j \geqslant 0$,

$$\sum_{\ell=-2^j}^{2^j - 1} \left| \hat{\psi}_2(2^j \omega - \ell) \right|^2 = 1 \quad \text{for } |\omega| \leqslant 1.$$

The function $\hat{\psi}_2$ is illustrated in Fig. 13a.

References

[1] J.P. Antoine, R. Murenzi, P. Vandergheynst, Directional wavelets revisited: Cauchy wavelets and symmetry detection in patterns, Appl. Comput. Harmon. Anal. 6 (1999) 314–345.
[2] A. Averbuch, R. Coifman, D. Donoho, M. Israeli, Y. Shkolnisky, Fast slant stack: A notion of Radon transform for data in a Cartesian grid which is rapidly computible, algebraically exact, geometrically faithful and invertible, SIAM J. Sci. Comput. (2007), in press.
[3] R.H. Bamberger, M.J.T. Smith, A filter bank for directional decomposition of images: Theory and design, IEEE Trans. Signal Process. 40 (1992) 882–893.
[4] P.J. Burt, E.H. Adelson, The Laplacian pyramid as a compact image code, IEEE Trans. Commun. 31 (4) (1983) 532–540.
[5] E.J. Candès, L. Demanet, D.L. Donoho, L. Ying, Fast discrete curvelet transforms, SIAM Multiscale Model. Simul. 5 (3) (2006) 861–899.
[6] E.J. Candès, D.L. Donoho, Ridgelets: A key to higher-dimensional intermittency? Philos. Trans. R. Soc. London A 357 (1999) 2495–2509.
[7] E.J. Candès, D.L. Donoho, New tight frames of curvelets and optimal representations of objects with piecewise C^2 singularities, Comm. Pure Appl. Math. 56 (2004) 216–266.
[8] R.R. Coifman, D.L. Donoho, Translation invariant de-noising, in: A. Antoniadis, G. Oppenheim (Eds.), Wavelets and Statistics, Springer-Verlag, New York, 1995, pp. 125–150.
[9] R.R. Coifman, F.G. Meyer, Brushlets: A tool for directional image analysis and image compression, Appl. Comput. Harmon. Anal. 5 (1997) 147–187.
[10] F. Colonna, G.R. Easley, Generalized discrete Radon transforms and their use in the ridgelet transform, J. Math. Imaging Vision 23 (2005) 145–165.
[11] A.L. Cunha, J. Zhou, M.N. Do, The non-subsampled contourlet transform: Theory, design, and applications, IEEE Trans. Image Process. 15 (2006) (2005) 3089–3101.
[12] M.N. Do, M. Vetterli, The contourlet transform: An efficient directional multiresolution image representation, IEEE Trans. Image Process. 14 (2005) 2091–2106.
[13] D.L. Donoho, Sparse components of images and optimal atomic decomposition, Constr. Approx. 17 (2001) 353–382.
[14] D.L. Donoho, M. Vetterli, R.A. DeVore, I. Daubechies, Data compression and harmonic analysis, IEEE Trans. Inform. Theory 44 (1998) 2435–2476.
[15] D.L. Donoho, I. Johnstone, Ideal spatial adaptation via wavelet shrinkage, Biometrica 81 (1994) 425–455.
[16] G.R. Easley, C.A. Berenstein, D.M. Healy Jr., Deconvolution in a ridgelet and curvelet domain, in: Independent Component Analysis, Wavelets, Unsupervised Smart Sensors, and Neural Networks III, Proc. of SPIE, vol. 5818, 2005.
[17] F. Falzon, S. Mallat, Analysis of low bit rate image transform coding, IEEE Trans. Image Process. 46 (4) (1998) 1027–1042.
[18] K. Guo, G. Kutyniok, D. Labate, Sparse multidimensional representations using anisotropic dilation and shear operators, in: G. Chen, M. Lai, (Eds.), Wavelets and Splines: Athens 2005, Proceedings of the International Conference on the Interactions between Wavelets and Splines, Athens, GA, May 16–19, 2005.
[19] K. Guo, D. Labate, Optimally sparse multidimensional representation using shearlets, SIAM J. Math. Anal. 39 (2007) 298–318.
[20] K. Guo, W. Lim, D. Labate, G. Weiss, E. Wilson, Wavelets with composite dilations, Electron. Res. Announc. Amer. Math. Soc. 10 (2004) 78–87.
[21] K. Guo, W. Lim, D. Labate, G. Weiss, E. Wilson, The theory of wavelets with composite dilations, in: C. Heil (Ed.), Harmonic Analysis and Applications, Birkhäuser, Boston, 2006, pp. 231–249.
[22] K. Guo, W. Lim, D. Labate, G. Weiss, E. Wilson, Wavelets with composite dilations and their MRA properties, Appl. Comput. Harmon. Anal. 20 (2006) 231–249.
[23] E. Hernández, G. Weiss, A First Course on Wavelets, CRC Press, Boca Raton, FL, 1996.
[24] L. Hörmander, The Analysis of Linear Partial Differential Operators. I. Distribution Theory and Fourier Analysis, Springer-Verlag, Berlin, 2003.

[25] N. Kingsbury, Complex wavelets for shift invariant analysis and filtering of signals, Appl. Comput. Harmon. Anal. 10 (2001) 234–253.
[26] I. Kryshtal, B. Robinson, G. Weiss, E. Wilson, Compactly supported wavelets with composite dilations, J. Geom. Anal. 17 (2006) 87–96.
[27] D. Labate, W. Lim, G. Kutyniok, G. Weiss, Sparse multidimensional representation using shearlets, in: Wavelets XI, San Diego, CA, 2005, in: SPIE Proc., vol. 5914, SPIE, Bellingham, WA, 2005.
[28] M. Lang, H. Guo, J.E. Odegard, C.S. Burrus, R.O. Wells Jr., Noise reduction using an undecimated discrete wavelet transform, Signal Process. Newslett. 3 (1996) 10–13.
[29] E. Le Pennec, S. Mallat, Sparse geometric image representations with bandelets, IEEE Trans. Image Process. 14 (2005) 423–438.
[30] W. Lim, Wavelets with composite dilations, Ph.D. thesis, Dept. Mathematics, Washington University in St. Louis, St. Louis, MO, 2006.
[31] Y. Lu, M.N. Do, Multidimensional directional filter banks and surfacelets, IEEE Trans. Image Process. 16 (2007) (2005) 918–931.
[32] S. Mallat, A Wavelet Tour of Signal Processing, Academic Press, San Diego, CA, 1998.
[33] D.D. Po, M.N. Do, Directional multiscale modeling of images using the contourlet transform, IEEE Trans. Image Process. 15 (2006) 1610–1620.
[34] L. Sendur, I.W. Selesnick, Bivariance shrinkage with local variance estimator, IEEE Signal Process. Lett. 9 (12) (2002) 438–441.
[35] J.L. Starck, E.J. Candès, D.L. Donoho, The curvelet transform for image denoising, IEEE Trans. Image Process. 11 (2002) 670–684.