Optimally Sparse Image Representations using Shearlets

Glenn R. Easley
System Planning Corporation,
1000 Wilson Blvd
Arlington, VA 22209 USA

Demetrio Labate
Department of Mathematics,
North Carolina State University,
Raleigh, NC 27695 USA

Wang–Q Lim
Department of Mathematics,
Washington University,
St. Louis, Missouri 63130 USA

***Abstract*— It is now widely acknowledged that traditional wavelets are not very effective in dealing with multidimensional signals containing distributed discontinuities. This paper presents a new discrete multiscale directional representation called the Discrete Shearlet Transform. This approach, which is based on the shearlet transform previously developed by the authors and their colaborators, 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. Numerical experiments demonstrate that the Discrete Shearlet Transform is very competitive in denoising applications both in terms of performance and computational efficiency.**

I. Introduction

The most useful feature of the wavelet transform is the ability to deal with signals containing isolated point singularities. This fact, together with the availability of fast discrete implementations, explains the spectacular success of wavelets in a variety of signal processing applications. Indeed, if a one-dimensional signal $s(t)$, smooth away from point discontinuities, 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 fact, it is significantly better than the corresponding Fourier approximation error [1], [2].

However, despite their optimal approximation properties for one-dimensional signals, traditional wavelet methods do not perform as well with multidimensional data. Indeed, wavelets are very efficient in dealing with isolated point singularities *only*. In higher dimensions, other types of singularities are usually present or even dominant. Images, for example, typically contain sharp transitions such as edges. Since edges interact extensively with the elements of the wavelet basis, "many" wavelet coefficients are needed to accurately represent these objects. A number of recent results have shown that a much more efficient representation of multidimensional data is obtained by better exploiting their geometric regularities. These various methods include *contourlets* [3], [4], *complex wavelets* [5] *brushlets* [6], *ridgelets* [7], *curvelets* [8], *bandelets* [9] and other schemes of "directional wavelets" [10].

The authors and their collaborators have recently introduced the *shearlet representation* [11], [12], which applies the framework of affine systems to capture very efficiently the geometry of multidimensional signals. As a result, this approach provides optimal approximation properties for a large class of two-dimensional images. The shearlet representation has a simple mathematical construction, extends naturally to higher dimensions and can be associated to a multiresolution analysis [12]. In addition, as we will show in this paper, this approach is amenable to a fast algorithmic implementation and is very competitive for image denoising.

The paper is organized as follows. In Section II we introduce a two-dimensional continuous transform called the Continuous Shearlet Trasform, which is well suited to locate discontinuities along edges. In Section III we show that the Discrete Shearlet Transform, obtained by discretizing the corresponding continuous transform, provides optimal representation for a large class of two-dimensional signals. In Section IV, we describe the algorithmic implementation of the Discrete Shearlet Transform and in Section V we describe some applications for image denoising.

II. The Continuous Shearlet Transform

An affine family generated by $\psi \in L^2(\mathbb{R})$ is a collection of functions of the form:

$$\{\psi_{a,t}(x) = a^{-1/2}\,\psi(a^{-1}x - t) : a > 0, t \in \mathbb{R}\}.$$

ψ is called a *continuous wavelet* if, for all $f \in L^2(\mathbb{R})$

$$f(x) = \int_0^\infty \int_{-\infty}^\infty \langle f, \psi_{a,t}\rangle\, \psi_{a,t}(x)\, dt\, \frac{da}{a}.$$

The *continuous wavelet transform* of f is:

$$\mathcal{W}f(a,t) = \langle f, \psi_{a,t}\rangle,$$

and the discrete wavelet transform is obtained by dicretizing $\mathcal{W}f(a,t)$ on an appropriate set [2].

The natural way of extending the theory of the continuous wavelet transform to higher dimensions is by considering the affine system

$$\{\psi_{M,t}(x) = |\det M|^{-1/2}\,\psi(M^{-1}x - t) : M \in G, t \in \mathbb{R}^n\},$$

where $\psi \in L^2(\mathbb{R}^n)$ and G is a subset of $GL_n(\mathbb{R})$, the group of invertible $n \times n$ matrices. Similarly to the one-dimensional case, ψ is called a *continuous wavelet* if, for all $f \in L^2(\mathbb{R}^n)$

$$f(x) = \int_G \int_{\mathbb{R}^n} \langle f, \psi_{M,t}\rangle\, \psi_{M,t}(x)\, dt\, d\lambda(M),$$

where $\lambda(M)$ is a measure on G, and

$$\mathcal{W}f(M,t) = \langle f, \psi_{M,t}\rangle$$

is the *continuous wavelet transform* of f [13].

The traditional way of discretizing the continuous wavelet transform of $f \in L^2(\mathbb{R}^n)$ replaces $M \in G, t \in \mathbb{R}^n$ with a discrete set $A^j, j \in \mathbb{Z}, k \in \mathbb{Z}^n$. However, starting from the continuous wavelet transform other types of discrete transforms can be deduced. Indeed, the shearlet transform, which will be described in the next section, is a obtained by discretizing the 2-dimensional continuous wavelet transform in a 'non-traditional' way.

For $\psi \in L^2(\mathbb{R}^2)$, consider the 2-dimensional affine system

$$\{\psi_{ast}(x) = |\det M_{as}|^{-\frac{1}{2}}\, \psi(M_{as}^{-1}x - t) : t \in \mathbb{R}^2,\, M_{as} \in \Gamma\}, \tag{1}$$

where Γ is the 2–parameter dilation group

$$\Gamma = \{M_{as} = \begin{pmatrix} a & \sqrt{a}\, s \\ 0 & \sqrt{a} \end{pmatrix} : (a, s) \in \mathbb{R}^+ \times \mathbb{R}\}.$$

We choose ψ such that

$$\hat{\psi}(\xi) = \hat{\psi}(\xi_1, \xi_2) = \hat{\psi}_1(\xi_1)\, \hat{\psi}_2(\frac{\xi_2}{\xi_1}), \tag{2}$$

where ψ_1 is a continuous wavelet for which $\hat{\psi}_1 \in C^\infty(\mathbb{R})$ with supp $\hat{\psi}_1 \subset [-2, 1/2] \cup [1/2, 2]$ and ψ_2 is chosen so that $\hat{\psi}_2 \in C^\infty(\mathbb{R})$, supp $\hat{\psi}_2 \subset [-1, 1]$, with $\hat{\psi}_2 > 0$ on (-1,1), and $\|\psi_2\| = 1$. Under these assumptions, ψ is a continuous wavelet [14] and for $a \in \mathbb{R}^+$, $s \in \mathbb{R}$, and $t \in \mathbb{R}^2$

$$Sf(a, s, t) = \langle f, \psi_{ast} \rangle$$

will be called the *continuous shearlet transform* of $f \in L^2(\mathbb{R})$.

The elements of the affine system, which we shall call *continuous shearlets*, are oriented waveforms whose orientation is controlled by the shear parameter s. They become increasingly elongated at fine scales (as $a \to 0$). We refer to [14] for more details.

III. Discrete Shearlet Transform

By sampling the Continuous Shearlet Transform $Sf(a, s, t)$ on an appropriate discrete set we obtain a discrete transform which is able to better deal with distributed discontinuities.

Observe that the matrix M_{as} can be factored as

$$\begin{pmatrix} a & \sqrt{a}\, s \\ 0 & \sqrt{a} \end{pmatrix} = \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a & 0 \\ 0 & \sqrt{a} \end{pmatrix}.$$

Thus, it will be "discretized" as $M_{j\ell} = B^\ell A^j$, where

$$B = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \quad A = \begin{pmatrix} 4 & 0 \\ 0 & 2 \end{pmatrix}$$

are the *shear matrix* and the *anisotropic dilation matrix*, respectively. Hence, the *discrete shearlets* are the functions of the form

$$\psi^{(0)}_{j,\ell,k}(x) = 2^{\frac{3j}{2}}\, \psi(B^\ell A^j x - k), \tag{3}$$

where

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

with $\hat{\psi}_1, \hat{\psi}_2 \in C^\infty(\widehat{\mathbb{R}})$, supp $\hat{\psi}_1 \subset [-\frac{1}{2}, -\frac{1}{16}] \cup [\frac{1}{16}, \frac{1}{2}]$ and supp $\hat{\psi}_2 \subset [-1, 1]$. In addition, we assume that

$$\sum_{j \geq 0} |\hat{\psi}_1(2^{-2j}\omega)|^2 = 1 \quad \text{for } |\omega| \geq \frac{1}{8}, \tag{4}$$

and, for each $j \geq 0$,

$$\sum_{\ell=-2^j}^{2^j-1} |\hat{\psi}_2(2^j\, \omega - \ell)|^2 = 1 \quad \text{for } |\omega| \leq 1. \tag{5}$$

From these assumptions it follows that the functions $\hat{\psi}^{(0)}_{j,\ell,k}$ (with $j \geq 0$, $-2^j \leq \ell \leq 2^j - 1$, $k \in \mathbb{Z}^2$) form a tiling of $\mathcal{D}_0 = \{(\xi_1, \xi_2) : |\xi_1| \geq \frac{1}{8}, |\frac{\xi_2}{\xi_1}| \leq 1\}$. This is illustrated in Figure 1(a). In very similar way, we construct a second set of discrete shearlets $\psi^{(1)}_{j,\ell,k}(x)$ such that the set $\{\hat{\psi}^{(1)}_{j,\ell,k} : j \geq 0, -2^j \leq \ell \leq 2^j - 1, k \in \mathbb{Z}^2\}$ is a tiling of $\mathcal{D}_1 = \{(\xi_1, \xi_2) : |\xi_2| \geq \frac{1}{8}, |\frac{\xi_1}{\xi_2}| \leq 1\}$ (see Figure 1(a)). Finally, let $\varphi \in L^2(\mathbb{R}^2)$ be such that the set $\{\varphi_k(x) = \varphi(x - k) : k \in \mathbb{Z}^2\}$ is a tight frame for $L^2([-\frac{1}{16}, \frac{1}{16}]^2)^\vee$. We deduce the following result.

Theorem 3.1 ([12]): The collection:

$$\{\varphi_k, \psi^{(d)}_{j,\ell,k} : j \geq 0, -2^j \leq \ell \leq 2^j - 1, k \in \mathbb{Z}^2, d = 0, 1\}$$

is a tight frame for $L^2(\mathbb{R}^2)$.

This indicates that the decomposition is invertible and the transformation is numerically well-conditioned.

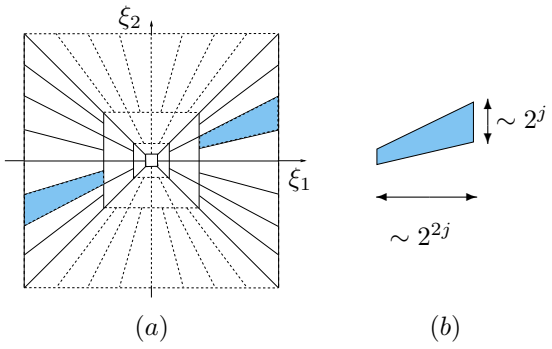

Fig. 1. *(a) The tiling of the frequency plane induced by the shearlets. The tiling of $\mathcal{D}_0$ is illustrated in solid line, the tiling of $\mathcal{D}_1$ is in dashed line. (b) The frequency support of a shearlet $\psi_{j,\ell,k}$ satisfies parabolic scaling.*

Details about this construction can be found in [12] and [15]. Let us summarize here the main mathematical properties of shearlets:

- Shearlets are *well localized*. 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 Figure 1(b)). 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:*
  Theorem ([11, Thm. 1.1]): Let f be C^2 away from piecewise C^2 curves, and let f_N^S be the approximation to f using the N largest coefficients in the shearlet expansion. Then

$$\|f - f_N^S\|_2^2 \le C\, N^{-2}\, (\log N)^3.$$

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.

IV. Algorithmic Implementation

It will be convenient to describe the collection of shearlets described above in a way which is more suitable to derive its numerical implementation. We introduce a collection of smooth window functions $W_{j,\ell}^{(d)}(\xi)$ localized on a pair of trapezoids, as illustrated in Figure 1(a), satisfying

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

The *discrete shearlet transform* of f

$$Sf(j,\ell,k) = \langle f, \psi_{j,\ell,k}^{(d)} \rangle$$

can be computed by

$$Sf(j,\ell,k) = \int_{\mathbb{R}^2} \hat{f}(\xi)\, \overline{V(2^{-2j}\,\xi)\, W_{j,\ell}^{(d)}(\xi)}\; e^{2\pi i \xi A_d^{-j} B_d^{-\ell} k}\, d\xi, \tag{6}$$

where $V(\xi_1,\xi_2) = \hat{\psi}_1(\xi_1)\, \chi_{\mathcal{D}_0}(\xi_1,\xi_2) + \hat{\psi}_1(\xi_2)\, \chi_{\mathcal{D}_1}(\xi_1,\xi_2)$.

A. A Frequency-Domain Implementation

Consider an $N \times N$ image given by $\{f[n_1,n_2]\}_{n_1,n_2=0}^{N-1,N-1}\}$. Its 2D Discrete Fourier Transform (DFT) is

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

where $-\frac{N}{2} \le k_1, k_2 < \frac{N}{2}$.

First, to compute

$$\hat{f}(\xi_1,\xi_2)\, \overline{V(2^{-2j}\xi_1, 2^{-2j}\xi_2)} \tag{7}$$

in the discrete domain, at the resolution level j, we apply the Laplacian pyramid algorithm [16]. This gives rise to the multiscale partition illustrated in Figure 1 by decomposing $f_a^{j-1}[n_1,n_2]$, $0 \le n_1, n_2 < N_{j-1}$, 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

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

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 $\widehat{f_d^j}(\xi_1,\xi_2)$.

In order to obtain the directional localization illustrated in Figure 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:

$$\begin{aligned} (u,v) &= (\xi_1, \tfrac{\xi_2}{\xi_1}) \quad \text{if } (\xi_1,\xi_2) \in \mathcal{D}_0 \\ (u,v) &= (\xi_2, \tfrac{\xi_1}{\xi_2}) \quad \text{if } (\xi_1,\xi_2) \in \mathcal{D}_1 \end{aligned}$$

Let us summarize the procedure illustrated by the scheme of Figure 2.

General Algorithm

Define f_a^0 be the given $N \times N$ image and set $N_0 = N$. For $j = 1, \ldots, L$, do the following:

1. Apply the Laplacian Pyramid scheme to decompose f_a^{j-1} into a low-pass $N_{j-1}/4 \times N_{j-1}/4$ image f_a^j and a high-pass $N_{j-1} \times N_{j-1}$image f_d^j.
2. Compute $\widehat{f_d^j}$ on a pseudo-polar grid and apply filtering $\hat{\psi}_2(\xi)$ along the angular direction to obtain $\widetilde{\widehat{f}}_d^j$.
3. Invert to obtain $\tilde{f}_d^j$.

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

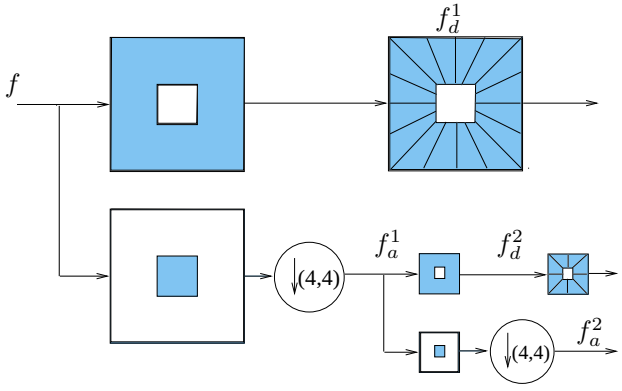

Fig. 2. *The figure illustrates the succession of Laplacian pyramid and directional filtering.*

Observe that, in this implementation, we have a large flexibility in the choice of the frequency window function W. In addition, there is a fast time-domain implementation of the windowing process. For more details about these issues, we refer to [15].

Figure 3 illustrates the two-level shearlet decomposition of the "Cameraman" image. The first-level decomposition generates 4 subbands, and the second-level decomposition generates

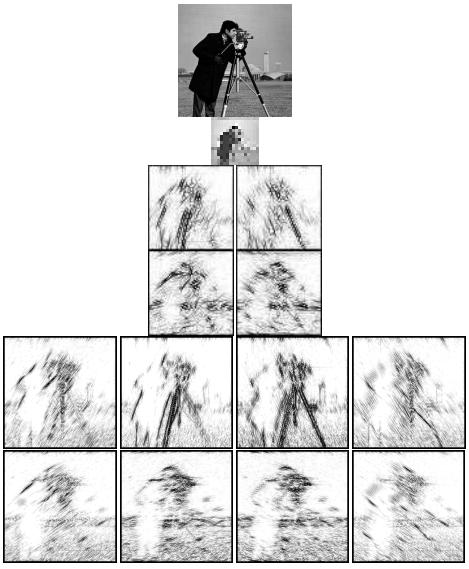

Fig. 3. The top image is the original Cameraman image. The image below the top image contains the approximate shearlet coefficients. Images of the detail shearlet coefficients are shown below this with an inverted grayscale for better presentation.

8 subbands, corresponding to the different directional bands illustrated by the scheme in Figure 2.

V. Summary and Experimental Results

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 $u = f+\epsilon$, where ϵ is Gaussian white noise with zero mean and standard deviation σ. We recover the image f from the noisy data u by computing an approximation $\tilde{f}$ of f obtained by applying a soft thresholding scheme in the subbands of the shearlet decomposition. We choose the threshold parameters $\tau_{i,j} = \sigma^2_{\epsilon_{i,j}}/\sigma^n_{i,j}$ as in [4], where $\sigma^n_{i,j}$ 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.

The particular form of the shearlet transform we tested was the nonsubsampled Laplacian pyramid transform with several different combinations of the shearing filters. In particular, we implemented the shearing on 4 scales of the Laplacian pyramid transform decomposition. The shearing filters of sizes 16×16, 16×16, 32×32, and 32×32 from finer to coarser were used with the number of shearing directions chosen to be 16, 16, 8, and 8. The shearing was done by using a Meyer wavelet window. In addition to the shearlet reconstruction, a slight post-filtering has been applied to the reconstructed estimate by using a stationary wavelet transform with a very small threshold parameter.

We tested the denoising schemes using the pieces of the "Peppers" and "Elaine" images for various standard deviation values of the noise (see Figures 4 and 5). For more competitive comparisons, we tested the scheme against the Curvelet based denoising scheme of [8] and the Nonsubsampled Contourlet Transform (NSCT) denoising scheme of [4] using 16, 16, 8, and 8 directions from finer to coarser scales.

TABLE I

	Noisy	Curvelet	NSCT	Shear
Peppers		PSNR (dB)		
$\sigma = 10$	28.11	32.25	33.75	33.79
$\sigma = 15$	24.58	31.36	32.43	32.54
$\sigma = 20$	22.09	30.60	31.43	31.57
Elaine		PSNR (dB)		
$\sigma = 10$	28.11	30.72	31.69	31.85
$\sigma = 15$	24.58	29.98	30.52	30.59
$\sigma = 20$	22.09	29.34	29.79	29.81

The performance of the shearlet approach relative to other transforms is shown in Table I. It shows that the shearlet algorithm consistently matches or outperforms all the algorithms mentioned above. It is noticeable that the shearlet transform results exhibit less Gibbs-type residual artifacts than the other denoising methods.

Further examples showing the good performance of the methods outlined in this work can be found in [15].

References

[1] D. L. Donoho, M. Vetterli, R. A. DeVore, I. Daubechies, Data compression and harmonic analysis, IEEE Trans. Inf. Th. 44 (1998) 2435–2476.

[2] S. Mallat, A Wavelet Tour of Signal Processing, Academic Press, San Diego, CA, 1998.

[3] M. N. Do, M. Vetterli, The contourlet transform: an efficient directional multiresolution image representation, IEEE Trans. Im. Proc. 14 (2005) 2091–2106.

[4] A. L. Cunha, J. Zhou, M. N. Do, The nonsubsampled contourlet transform: Theory, design, and applications, in press.

[5] N. Kingsbury, Complex wavelets for shift invariant analysis and filtering of signals, Appl. Computat. Harmon. Anal. 10 (2001) 234-253.

[6] R. R. Coifman, F. G. Meyer, Brushlets: a tool for directional image analysis and image compression, Appl. Comp. Harmon. Anal. 5 (1997) 147–187.

[7] E. J. Candès, D. L. Donoho, Ridgelets: a key to higher–dimensional intermittency?, Phil. Trans. Royal Soc. London A 357 (1999), 2495–2509.

[8] J. L. Starck, E. J. Candès, D. L. Donoho, The curvelet transform for image denoising, IEEE Trans. Im. Proc., 11 (2002) 670-684.

[9] E. Le Pennec, and S. Mallat, Sparse geometric image representations with bandelets, IEEE Trans. Image Process. 14 (2005) 423–438.

[10] J. P. Antoine, R. Murenzi, P. Vandergheynst, Directional wavelets revisited: Cauchy wavelets and symmetry detection in patterns, Appl. Computat. Harmon. Anal. 6 (1999) 314-345.

[11] K. Guo, D. Labate, Optimally Sparse Multidimensional Representation using Shearlets, preprint 2006.

[12] K. Guo, W-Q. Lim, D. Labate, G. Weiss and E. Wilson, Wavelets with composite dilations and their MRA properties, Appl. Computat. Harmon. Anal. 20 (2006) 231–249.

[13] G. Weiss, and E. Wilson, The mathematical theory of wavelets, Proceeding of the NATO–ASI Meeting. Harmonic Analysis 2000. A Celebration. Kluwer Publisher, 2001.

[14] G. Kutyniok, and D. Labate, Resolution of the wavefront set using shearlets, preprint (2006).

[15] G. R. Easley, D. Labate, W.Q Lim, Sparse Directional Image Representations using the Discrete Shearlet Transform, preprint (2006).

[16] P. J. Burt, E. H. Adelson, The Laplacian pyramid as a compact image code, IEEE Trans. Commun. 31(4)(1983) 532-540.

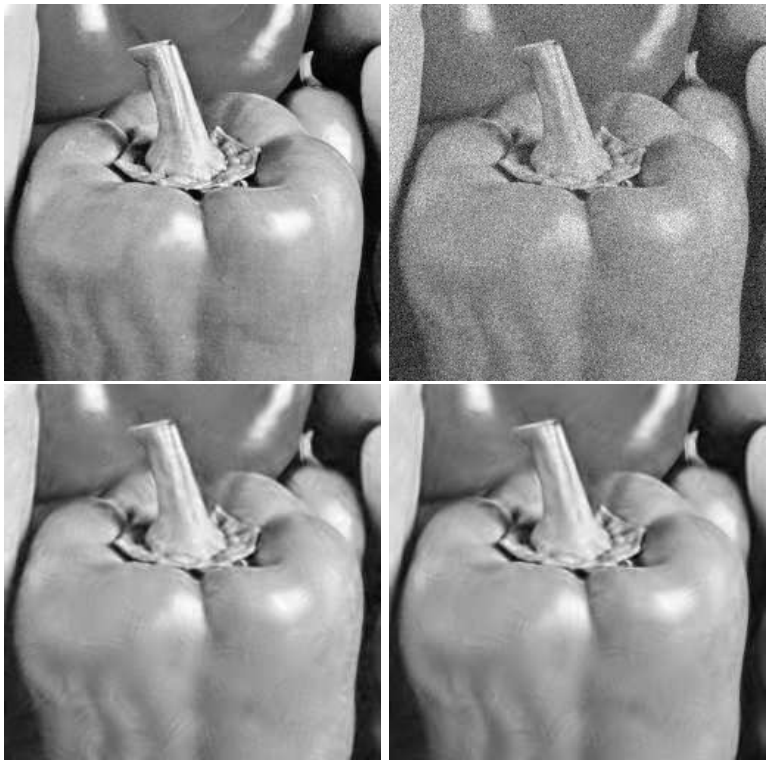

Fig. 4. Image denoising results of a piece of the "Peppers" with a standard diviation of 20. From top left, clockwise: Original image, noisy image (PSNR= 22.09 dB), Shearlet transform (PSNR=31.57 dB), and NSCT (PSNR= 31.43 dB)

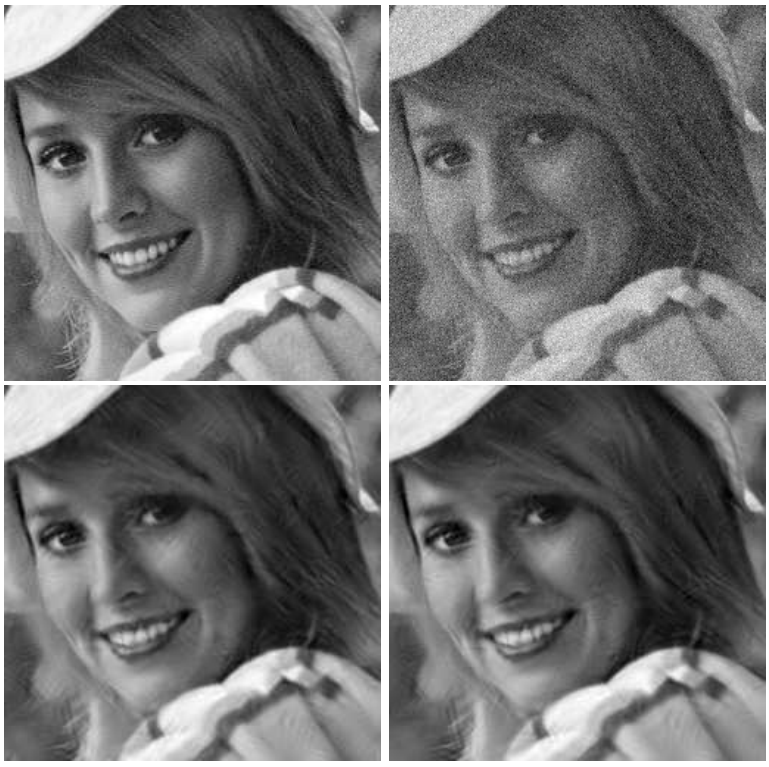

Fig. 5. Image denoising results of a piece of the "Elaine" with a standard diviation of 20. From top left, clockwise: Original image, noisy image (PSNR= 22.09 dB), Shearlet transform (PSNR=29.81 dB), and NSCT (PSNR= 29.79 dB)