

Image Processing using Shearlets

Glenn R. Easley and Demetrio Labate

Abstract Since shearlets provide nearly optimally sparse representations for a large class of functions that are useful to model natural images, many image processing methods benefit from their use. In particular, the error rates of data estimation from noise are highly dependent on the sparsity properties of the representation, so that many successful applications of shearlets center around restoration tasks such as denoising and inverse problems. Other imaging problems, where also the application of the shearlet representation turns out to be very beneficial, include image enhancement, image separation, edge detection, and estimation of the geometric features of an object.

Key words: curvelets, deconvolution, denoising, edge detection, geometric separation, image processing, shearlets, sparsity, wavelets, video denoising

1 Introduction

Shearlets were introduced with the expressed intent to provide a highly efficient representation of images with edges. In fact, the elements of the shearlet representation form a collection of well-localized waveforms, ranging at various locations, scales and orientations, and with highly anisotropic shapes. This makes the shearlet representation particularly well adapted at representing the edges and the other anisotropic objects which are the dominant features in typical images. These properties have been theoretically quantified through the notion of sparse shearlet approximations and the shearlet analysis of singularities (see Chapters 2 and 3 of

Glenn R. Easley
System Planning Corporation, Arlington, VA 22201, USA, e-mail: geasley@sysplan.com

Demetrio Labate
University of Houston, Department of Mathematics, Houston, TX 77204, USA,
e-mail: dlabate@math.uh.edu

this volume). As will be described below, these properties have direct and important implications for the efficient encoding and processing of discrete data. This is demonstrated by an increasing number of very competitive numerical applications of the shearlet transform to the analysis and processing of images and other multi-dimensional data.

In this chapter, we provide an overview of the most relevant imaging applications of the shearlet approach. For reason of space, we will focus on the general principles used in the development of the algorithms and their significance, rather than on the technical details of the implementations, which can be found in the original papers. Since this is a very active area of investigation, it is understood that improved and newer shearlet imaging applications are currently being developed and, as a consequence, this chapter can only provide a retrospective view on the field. In particular, this chapter will describe the application of the shearlet representation to problems of image denoising, image enhancement, inverse problems, edge analysis and detection, and image separation.

2 Image denoising

The significance of sparsity for data restoration is well understood and has been addressed in seminal papers such as [20, 27]. Indeed, consider the classical problem of recovering a function $f \in L^2(\mathbb{R})$ from noisy data y , that is, of recovering f from the observation

$$y = f + n,$$

where n is Gaussian white noise¹ with standard deviation σ . We illustrate this problem in dimension $D = 1$, but it generalizes naturally to higher dimensions.

The problem of interest is to optimize the estimation \tilde{f} of f by minimizing the estimation error, usually measured by the L^2 norm $\|f - \tilde{f}\|$. Hence, the *risk* of the estimator \tilde{f} is given by the Mean Square Error (MSE)

$$E\|f - \tilde{f}\|^2,$$

where the expectation is calculated with respect to the probability distribution of the noise n . It is clear that the risk depends on f , so that the worst behaviour of the estimator is obtained by considering the supremum over all f in a certain class \mathcal{F} , that is,

$$\sup_{f \in \mathcal{F}} E\|f - \tilde{f}\|^2.$$

The *Minimax MSE* is defined as

¹ In practice, it is not always accurate to assume that the noise found in applications is Gaussian white noise. However, this assumption is usually needed to make the theory tractable. In particular, it is a standard assumption in the theory of wavelet thresholding and wavelets shrinkage which is discussed below.

$$\inf_{\tilde{f}} \sup_{f \in \mathcal{F}} E \|f - \tilde{f}\|^2,$$

where all measurable estimation procedures are allowed in the infimum. It is a remarkable result and an important application of the sparsity properties of wavelets that, for uniformly regular and piecewise regular 1D signals, a nearly minimax MSE is achieved using a very simple wavelet estimator known as *wavelet thresholding* [22, 23]. The computation of the wavelet threshold estimator can be briefly described as follows. Let $\{\psi_{j,m}\}$ be a wavelet basis for $L^2(\mathbb{R})$. Then the noisy function y can be expanded as

$$y = \sum_{j,k} \langle y, \psi_{j,m} \rangle \psi_{j,m},$$

with convergence in the L^2 norm. The wavelet *hard thresholding algorithm* consists in setting to zero the wavelet coefficients $\langle y, \psi_{j,m} \rangle$ whose absolute values fall below a certain threshold T . The value of T depends on the standard deviation of the noise σ . Hence, to summarize, the wavelet hard thresholding algorithm consists of the following steps:

1. Compute the wavelet coefficients $\langle y, \psi_{j,m} \rangle$ of y .
2. Determine the threshold value $T(\sigma)$, where σ is estimated from y .
3. Remove (zero out) the wavelet coefficients whose absolute values are smaller than $T(\sigma)$.
4. Compute the estimator \tilde{f} as $\tilde{f} = \sum_{j,k} c_{j,k} \psi_{j,m}$, where

$$\begin{aligned} c_{j,k} &= \langle y, \psi_{j,m} \rangle, & \text{if } |\langle y, \psi_{j,m} \rangle| > T, \\ c_{j,k} &= 0, & \text{otherwise.} \end{aligned}$$

An alternative thresholding approach is to use a *soft thresholding* algorithm, where the coefficients are modified by the *shrinkage function*: $\text{shr}(c) = \text{sgn}(c) \max(|c| - T, 0)$. Unlike the hard thresholding that is an 'all or nothing procedure' (values above the threshold are kept, values below it are deleted), the soft thresholding function produces a smooth transitions between the original and the deleted values, where values slightly below the threshold are not removed but attenuated. In practice, the main challenge is to find an appropriate value of T , and several strategies have been proposed in the literature [59]. For example, the VisuShrink algorithm [23, 24] uses the universal threshold $T = \sigma \sqrt{2 \log M}$, M being the size of the data, and is asymptotically near minimax within the class of Besov spaces. Another classical approach, called BayesShrink [12], uses a different threshold value $T_j = \frac{\sigma_j^2}{\sigma_j}$ for each resolution level j , where σ_j is the standard deviation of the data at the resolution level j .

As discussed in the previous chapters, wavelets are non-optimal when dealing with piecewise regular multivariable functions, and this implies that wavelet thresholding does not provide a minimax MSE in this situation. Consider, in particular, the class of cartoon-like images $\mathcal{E}^2(\mathbb{R}^2)$. For $f \in \mathcal{E}^2(\mathbb{R}^2)$, define $|c(f)^W|_{(N)}$ to be the N -th largest entry in the sequence of wavelet coefficients of f given by

$\{|c(f)_\mu^W| : c(f)_\mu^W = \langle f, \psi_\mu \rangle\}$ where $\{\psi_\mu\}$ is a 2D wavelet basis. Since

$$\sup_{f \in \mathcal{C}^2(\mathbb{R}^2)} |c(f)_\mu^W|_{(N)} \leq CN^{-1},$$

it follows that the N -term wavelet estimator \tilde{f}_N satisfies

$$\|f - \tilde{f}_N\|^2 \leq \sum_{m>N} |c(f)_\mu^W|_{(m)}^2 \leq CN^{-1}.$$

This implies that the Mean Square Error (MSE) of the wavelet thresholding estimator satisfies

$$\sup_{f \in \mathcal{E}} E\|f - \tilde{f}\|^2 \asymp \sigma, \quad \text{as } \sigma \rightarrow 0,$$

where σ is the noise level, as indicated above. By contrast, let $|c(f)_\mu^S|_{(N)}$ be the N -th largest entry in the sequence of shearlet coefficients of f given by $\{|c(f)_\mu^S| : c(f)_\mu^S = \langle f, s_\mu \rangle\}$ where $\{s_\mu\}$ is a Parseval frame of shearlets. Then a basic result from [43] shows that

$$\sup_{f \in \mathcal{E}} |c(f)_\mu^S|_{(N)} \leq CN^{-3/2} (\log N)^{3/2}.$$

Ignoring the log factor, this gives that the N -term shearlet estimator \tilde{f}_N satisfies

$$\|f - \tilde{f}_N\|^2 \leq \sum_{m>N} |c(f)_\mu^S|_{(m)}^2 \leq CN^{-2}.$$

This implies that a denoising strategy based on the thresholding of the shearlet coefficients yields an estimator \tilde{f} of f whose Mean Square Error (MSE) satisfies (essentially) the minimax MSE

$$\sup_{f \in \mathcal{E}} E\|f - \tilde{f}\|^2 \asymp \sigma^{4/3}, \quad \text{as } \sigma \rightarrow 0,$$

This shows that a denoising estimator based on shearlet thresholding has the ability to achieve a minimax MSE for images with edges. Notice that the same type of theoretical behaviour is achieved using curvelets [77, 6].

In the following, we will show some numerical demonstrations to illustrate that indeed a denoising algorithm based on shearlet thresholding outperforms a similar wavelet-based approach. Before presenting these results, let us briefly recall the construction of the Discrete Shearlet Transform, originally introduced in [32].

2.1 Discrete Shearlet Transform

We start by re-writing the cone-based shearlet system, defined in Chapter 1 of this volume, in a form more suitable to its digital implementation. As usual, it is more convenient to work in the Fourier domain. For additional emphasis, in the following we will use the notation $\widehat{\mathbb{R}}^2$ when referring to the plane in the Fourier domain, while we will use the standard symbol \mathbb{R}^2 to denote the plane in the space domain.

Set $\mathcal{D}_0 = \{(\xi_1, \xi_2) \in \widehat{\mathbb{R}}^2 : |\xi_1| \geq \frac{1}{8}, |\frac{\xi_2}{\xi_1}| \leq 1\}$ and $\mathcal{D}_1 = \{(\xi_1, \xi_2) \in \widehat{\mathbb{R}}^2 : |\xi_2| \geq \frac{1}{8}, |\frac{\xi_1}{\xi_2}| \leq 1\}$. Given a smooth function $\hat{\psi}_2$, with support in $[-1, 1]$, we define

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

and

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

for $\xi = (\xi_1, \xi_2) \in \widehat{\mathbb{R}}^2$, $j \geq 0$, and $\ell = -2^j, \dots, 2^j - 1$. This notation allows us to write the elements of the cone-based shearlet system, in the Fourier domain, as

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

where $d \in \{0, 1\}$, $V(\xi_1, \xi_2) = \hat{\psi}_1(\xi_1) \chi_{\mathcal{D}_0}(\xi_1, \xi_2) + \hat{\psi}_1(\xi_2) \chi_{\mathcal{D}_1}(\xi_1, \xi_2)$ and $\hat{\psi}_1$ is the Meyer-type wavelet associated with the classical shearlet. Hence, the *shearlet transform* of $f \in L^2(\mathbb{R}^2)$ can be expressed as

$$\langle f, \psi_{j,\ell,m}^{(d)} \rangle = 2^{\frac{3j}{2}} \int_{\mathbb{R}^2} \hat{f}(\xi) \overline{V(2^{-2j} \xi) W_{j,\ell}^{(d)}(\xi)} e^{2\pi i \xi A_d^{-j} B_d^{-\ell} m} d\xi. \quad (1)$$

To formulate the implementation in the finite domain setting, we consider $\ell^2(\mathbb{Z}_N^2)$ as the discrete analogue of $L^2(\mathbb{R}^2)$. Given an image $f \in \ell^2(\mathbb{Z}_N^2)$, the Discrete Fourier Transform (DFT) is defined as

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

The product $\hat{f}(\xi_1, \xi_2) \overline{V(2^{-2j} \xi_1, 2^{-2j} \xi_2)}$ is found analogously in the DFT domain as the product of the DFT of f with the discretization of the *filter functions* $\overline{V(2^j \cdot)}$, $j \geq 0$. Notice that these functions are associated to specific regions of the frequency

plane, roughly near $|\xi_1| \approx 2^{2j}$ or $|\xi_2| \approx 2^{2j}$. In the space domain, this produces a decomposition of f , at various scales j , in terms of elements of the form $f^j(n_1, n_2) = f * v_j(n_1, n_2)$, where v_j corresponds to the function $\overline{V(2^j \cdot)}$ in the Fourier domain. This can be implemented using the Laplacian pyramid filter [5].

In order to obtain the directional localization, \hat{f}^j is resampled onto a pseudo-polar grid, and a one-dimensional band-pass filter is applied. The pseudo-polar grid is parametrized by lines going through the origin and their slopes. Specifically, the pseudo-polar coordinates $(u, p) \in \mathbb{R}^2$ match the following assignments:

$$\begin{aligned} (u, p) &= \left(\xi_1, \frac{\xi_2}{\xi_1} \right) & \text{if } (\xi_1, \xi_2) \in \mathcal{D}_0 \\ (u, p) &= \left(\xi_2, \frac{\xi_1}{\xi_2} \right) & \text{if } (\xi_1, \xi_2) \in \mathcal{D}_1 \end{aligned}$$

The resampled \hat{f}^j is denoted as F_j , so that

$$\hat{f}(\xi_1, \xi_2) \overline{V(2^{-2j}\xi_1, 2^{-2j}\xi_2)} W_{j,\ell}^{(d)}(\xi_1, \xi_2) = F_j(u, p) \overline{W^{(d)}(2^j p - \ell)}. \quad (2)$$

This resampling in the DFT can be done by direct re-assignment or by using the Pseudo-polar DFT [2].

Given the one-dimensional DFT defined as

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

we denote $\{w_{j,\ell}^{(d)}(n) : n \in \mathbb{Z}_N\}$ to be the sequence of values such that $\mathcal{F}_1\left(w_{j,\ell}^{(d)}(n) = \overline{W^{(d)}(2^j n - \ell)}\right)$. For a fixed $n_1 \in \mathbb{Z}_N$, we then have

$$\mathcal{F}_1\left(\mathcal{F}_1^{-1}\left(F_j(n_1, n_2)\right) * w_{j,\ell}^{(d)}(n_2)\right) = F_j(n_1, n_2) \mathcal{F}_1\left(w_{j,\ell}^{(d)}(n_2)\right), \quad (3)$$

where $*$ denotes the one-dimensional convolution along the n_2 axis. Equation (3) summarizes the algorithmic implementation for computing the discrete samples of $F_j(u, p) \overline{W^{(d)}(2^j p - \ell)}$.

The shearlet coefficients $\langle f, \psi_{j,\ell,m}^{(d)} \rangle$, given by (1), are now formally obtained by inverting the Pseudo-polar Fourier transform of expression (2). This can be either implemented by computing the inverse Pseudo-polar DFT or by directly re-assembling the Cartesian sampled values of (3) and apply the inverse two-dimensional DFT. Using the Fast Fourier Transform (FFT) to implement the DFT, the discrete shearlet transform algorithm runs in $O(N^2 \log N)$ operations. Note that the direct conversion between the Cartesian to pseudo-polar can be pre-conditioned so the operation has a condition number of 1 as explained in [32] and the operation will be L^2 norm preserving (cf. [16] for details). Since no performance improvement was noticed with this adjustment, this pre-conditioning was avoided at the expense

of having a faster inversion, which is just a summation, since it was observed to have many advantages for this formulation.

An illustration of the shearlet decomposition produced by this implementation is given in Figure 1, which shows a 2-level subband decomposition, that is, a decomposition where the scale parameter j is ranging over $j = 0, 1$. Recall that, in the language of image processing, a *subband decomposition* is a decomposition of an image into components associated with different regions of the frequency plane $\widehat{\mathbb{R}}^2$. As the figures shows, there are 4 subband terms corresponding to $j = 0$ and 8 subband terms corresponding to $j = 1$, consistent with the fact that the directional parameter ℓ takes values in $\{-2^j, \dots, 2^j - 1\}$ for each of the cone regions $d = 0$ and $d = 1$. This corresponds precisely to the illustration of the shearlet decomposition into directional subbands given in Chapter 1 of this volume. A Matlab toolbox for this numerical implementation of the Discrete Shearlet Transform is available from <http://www.math.uh.edu/~dlabate>.

An alternative technique to implement the discrete shearlet transform as an application of M filters was presented in [33]. In this case, filters v_j and $w_{j,\ell}^{(d)}$ are found so that $\langle f, \psi_{j,\ell,m}^{(d)} \rangle$ can be computed as

$$f * (v_j * w_{j,\ell}^{(d)})(m) \triangleq f * g_{j,\ell}^{(d)}[m],$$

where $g_{j,\ell}^{(d)} = v_j * w_{j,\ell}^{(d)}$ are the directionally-oriented filters. When the filters $g_{j,\ell}$ are chosen to have significantly smaller support sizes than N as explained in [32], the filter bank implementation is even faster than $O(N^2 \log N)$. It is also a formulation that is easily parallelizable, in the sense that the different directional components of the image can be processed in parallel.

Besides these implementations, it is useful to recall that a reduced-redundancy implementation of the discrete shearlet transform was presented in [42] and a critically sampled version of the discrete shearlet transform was presented [30]. Finally, a compactly supported version of the discrete shearlet transform is discussed in [55]. We refer to Chapter 7 of this volume for additional detail about the digital implementations of the discrete shearlet transform.

To simplify the notation, in the applications which will be presented below, the superscript (d) will be suppressed and the distinction between $d = 0$ and $d = 1$ will be absorbed by re-indexing the parameter ℓ so that the cardinality is doubled.

2.2 Shearlet Thresholding

In this section, we present the first application of the discrete shearlet transform to the problem of image denoising. This approach can be viewed as a direct adaptation of the classical wavelet thresholding described above.

Suppose that an image f is corrupted by white Gaussian noise with standard deviation σ . Using the discrete shearlet transform, an estimator \tilde{f} of f can be com-



Fig. 1 An illustration of the subband decomposition obtained using the discrete shearlet transform. The top image is the original *Barbara* image. The image below the top image is of the coarse-scale reconstruction. Images of the subband reconstructions for levels $j = 0$ and $j = 1$ are given below with an inverted grayscale for presentation purposes. As explained in the text, there are 4 directional subbands corresponding to $j = 0$ and 8 directional subbands corresponding to $j = 1$.

puted by using a thresholding procedure which follows essentially the same ideas of the wavelet thresholding algorithm described in Sec. 2. In particular, for the choice

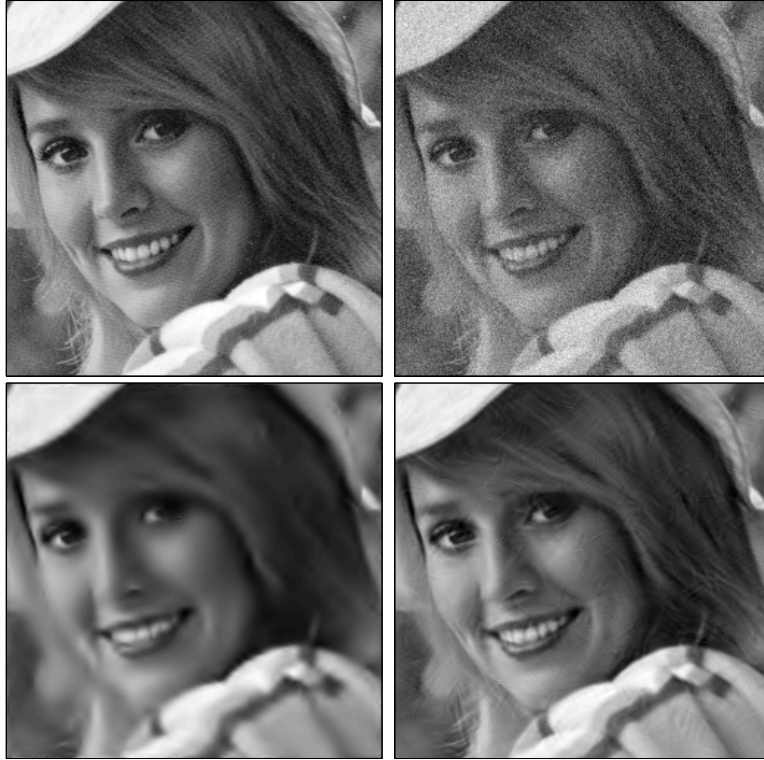


Fig. 2 Image denoising results of a piece of the *Elaine*. From top left, clockwise: Original image, noisy image (SNR= 10.46 dB), shearlet transform (SNR=16.47 dB), and wavelet transform (SNR= 14.00 dB)

of the threshold parameter it was found in [32] that an excellent performance is achieved by adapting the BayesShrink algorithm. That is, the threshold is chosen to be $T_{j,\ell} = \frac{\sigma_{j,\ell}^2}{\sigma_{j,\ell,m}}$, where $\sigma_{j,\ell}$ is the standard deviation of the noise for the subband indexed by the scaling parameter j and the directional parameter ℓ , and $\sigma_{j,\ell,m}$ denotes the standard deviation of the m -th coefficient of the image at scale j and direction ℓ .

A numerical demonstration of this shearlet-based denoising approach is given in Figure 2, where this method is compared against a similar scheme based on a nonsubsampling wavelet transform. To assess the performance of the algorithm, we have used the standard *signal-to-noise* ratio to measure how much noise is present before and after the estimate is made. Recall that the signal-to-noise ratio (SNR) is given by

$$SNR(f, f_{\text{est}}) = 10 \log_{10} \left[\frac{\text{var}(f)}{\text{mean}(f - \tilde{f})} \right],$$

where $\text{var}(f)$ is the variance of the image, and is measured in decibels.

We refer to [32] for additional numerical tests and details about this version of the shearlet thresholding algorithm.

2.3 Denoising using Shearlet-based Total Variation Regularization

An alternative approach for the application of shearlets to problems of image denoising consists in combining the ideas of wavelet thresholding described above with other classical methods for denoising.

Indeed, there is another very successful philosophy to image denoising which is based on the theory of the partial differential equations and variational methods, such as diffusion equations and total variation (TV) minimization (cf. [10, 73, 88]). Intuitively, the idea of diffusion equations is to model a noisy image as a function \tilde{f} on $\Omega \subset \mathbb{R}^2$ and to compute its denoised version as the solution of a suitable diffusion process (isotropic and anisotropic) with \tilde{f} as initial condition. It is clear that this produces an image which is more “regular” than the original one. Alternatively, one can produce a similar regularization process by minimizing an energy functional of the form

$$E(f; \lambda, \tilde{f}) = \frac{\lambda}{2} \int_{\Omega} (f - \tilde{f})^2 dx dy + P(f),$$

where the first term, called *fidelity term*, encourages the similarity between \tilde{f} and its denoised version f , and the second term $P(f, \nabla f)$, called *penalty term*, controls the regularity of the solution. Indeed, it is known that there is a strong relation between regularization methods and diffusion filters [74].

In particular, let us consider a classical version of a regularization method based on *Total Variation (TV) regularization*, which consists in minimizing the functional

$$\int_{\Omega} \phi(\|\nabla f\|) dx dy + \frac{\lambda}{2} \int_{\Omega} (f - \tilde{f})^2 dx dy,$$

where $\phi \in C^2(\mathbb{R})$ is an even regularization function (cf. [4]). In the above expression, the penalty term involves the *total variation* of f , which, for a function $f \in W^{1,1}(\Omega)$, is defined as

$$TV(f) = \int_{\Omega} \|\nabla f\| dA,$$

where $\nabla f = \left(\frac{\partial u}{\partial x_1}, \frac{\partial u}{\partial x_2} \right)$ and $\|\cdot\|$ is the standard Euclidean norm. That is, the penalty term ensures that the solution of the regularization method minimizes the total variation of f (cf. [62] for a discussion of the role of TV in imaging applications). The minimizer can be found by solving the equation

$$\frac{\partial f}{\partial t} = \nabla \cdot \left(\frac{\phi'(\|\nabla f\|)}{\|\nabla f\|} \nabla f \right) - \lambda(f - f_0),$$

subjected to the von Neumann boundary condition. In the case when $\lambda = 0$ and $\lim_{x \rightarrow \infty} \phi'(x)/x = 0$, this equation is considered a special case of the Perona and Malik diffusion equation (cf. [71]):

$$\frac{\partial f}{\partial t} = \nabla \cdot (\rho(\|\nabla f\|)\nabla f),$$

where $\rho(x) = \phi'(x)/x$. There are several other aspects of these problems which go beyond the space limitations of this chapter, and we refer the interested reader to the references mentioned at the beginning of this section.

The regularization methods described above can be very effective in image denoising, and generally provide superior denoising performance especially when applied to images with negligible texture and fine-scale features. Yet a drawback is that they can result in estimates that are reminiscent of oil-paintings, with loss of important detail when applied to images that contain complex textures and shading. To improve upon these methods, combinations of such techniques with sparse representations have recently been proposed (e.g. [7, 15, 28, 58, 80, 89]). A similar combination has been proposed using shearlets in [31].

The idea of the shearlet-based Total Variation (TV) Regularization approach is rather simple yet very effective. Assuming a denoised estimate is found by thresholding a shearlet representation (using the method from Sec. 2.2), let P_S be the projection operator that retains the non-threshold shearlet coefficients of f . The shearlet-based TV method is then described as essentially solving

$$\frac{\partial f}{\partial t} = \nabla \cdot \left(\frac{\phi'(\|\nabla P_S(f)\|)}{\|\nabla P_S(f)\|} \nabla P_S(f) \right) - \lambda_{x,y}(f - \tilde{f})$$

with the boundary condition $\frac{\partial f}{\partial n} = 0$ on $\partial\Omega$ and the initial condition $f(x, y, 0) = \tilde{f}(x, y)$ for $x, y \in \Omega$. Note that the fidelity parameter $\lambda_{x,y}$ is spatially varying. It is based on a measure of local variances that is updated after a number of iterations L or progressions of artificial time steps (see [40] for more details).

Another diffusion variant based on shearlets has been to solve

$$\frac{\partial f}{\partial t} = \nabla \cdot (\rho(\|\nabla P_S f\|)\nabla P_S f)$$

with the Neumann boundary condition $\frac{\partial f}{\partial n} = 0$ on $\partial\Omega$ and the initial condition $f(x, y, 0) = f_0(x, y)$ for $x, y \in \Omega$.

Illustrations of these techniques, including a comparison with standard TV, are done using an image of flowers. Close-ups of the results are shown in Figure 3.

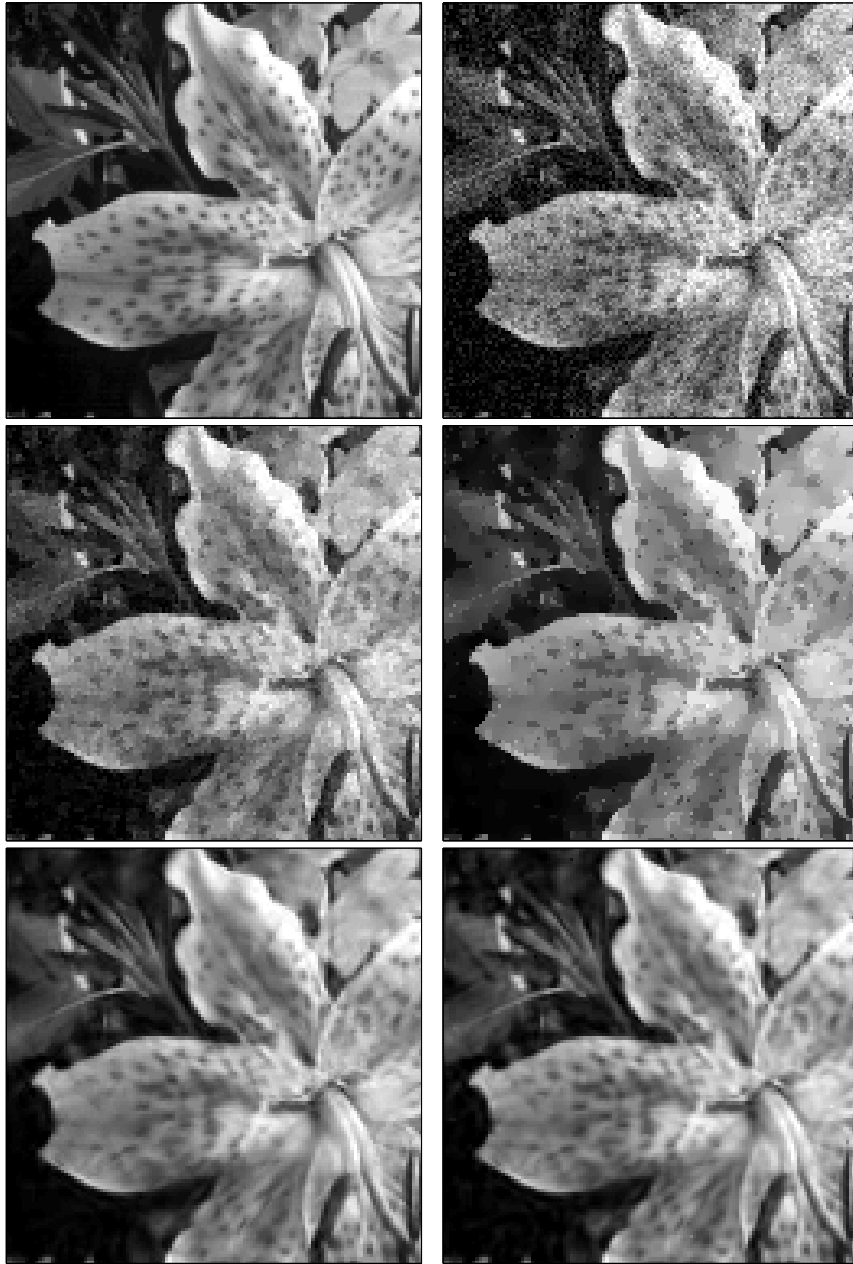


Fig. 3 Detail images of experimental results. From the top, clockwise: Original image, noisy image (SNR=11.11 dB), diffusion based estimate using 53 iterations (SNR=14.78 dB), shearlet based diffusion estimate using 6 iterations (SNR=16.15 dB), shearlet based TV estimate using 2 iterations ($L = 7$, SNR=16.29 dB), TV based estimate using 113 iterations (SNR=14.52 dB).

2.4 Complex-Valued Denoising

Another variant of the shearlet denoising algorithm was developed to deal with the problem of reducing complex-valued noise [67]. This problem arises in synthetic aperture radar (SAR) interferometry where interferometric phase noise reduction is a main challenge.

Recall that an SAR image is a complex-valued two-dimensional array and is often displayed in terms of its magnitude without any phase information. Interferograms are obtained by multiplying a SAR image by the complex conjugate of a second SAR image obtained from a slightly different location. These interferograms contain information about topographic height and are used to produce digital elevation maps (DEM). A typical problem is that the complex-valued noise in these phase estimates cause errors with the phase unwrapping needed for the height information to be formed.

The n -looks complex image is defined as

$$f = \frac{1}{n} \sum_{k=1}^n f_1(k) f_2^*(k) = |f| e^{j\psi}, \quad (4)$$

where f_1 and f_2 are a pair of 1-look complex SAR images. The phase quality depends on the amplitude of the correlation coefficient and is given as

$$\rho = \frac{E[f_1 f_2^*]}{\sqrt{E[|f_1|^2] E[|f_2|^2]}} = |\rho| e^{j\theta}, \quad (5)$$

where $|\rho|$ is the coherence and θ is the phase of the complex correlation coefficient. By using an appropriate phase noise model, a shearlet coefficient shrinkage method can be derived which adapts the one presented in Sec. 2.2. Figure 2.4 gives an illustration of this method and compares it to a wavelet-based method. In this example a single look image is given. The difference between the ideal phase image and the estimate is given by counting the number of residues.

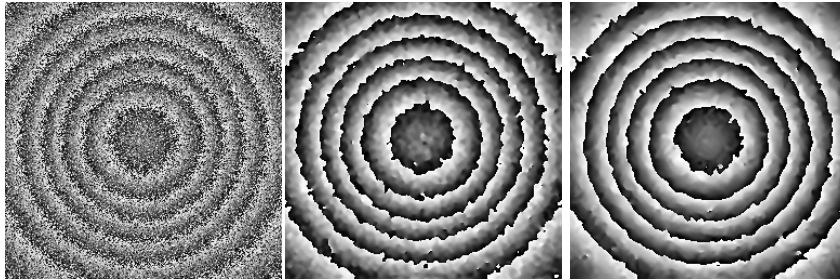


Fig. 4 Noisy interferometric phase filtering methods. From left to right: Noisy interferogram with coherence $|\rho| = 0.5$ (number of residues is 14119), wavelet-based estimate (number of residues is 80), shearlet-based estimate (number of residues is 20).

2.5 Other Shearlet-based Denoising Techniques

Additional methods of image denoising based on shearlets were recently presented in the literature [9, 13, 14, 18, 46, 82, 84, 92, 93] but discussing those in detail would go beyond the scope of this chapter. Many of these methods use variants of the shearlet shrinkage strategy described above, or they introduce some form of adaptivity in the thresholding.

3 Inverse Problems

In many scientific and industrial applications, the objects or the features of most interest cannot be observed directly, but must be inferred from other observable quantities. The aim of inverse problems is to reconstruct the cause for such observed variables. An especially important class of inverse problems, for example, concerns the determination of the structural properties of an object from measurements of the absorbed or scattered radiation taken at an array of sensors, which occurs in applications such as Computerized Tomography (CT) or Synthetic Aperture Radar (SAR). Another example is the removal of image degradation due to optical distortion or motion blur.

In most of these problems, the relationship between the observed data y and the feature of interest f is approximately linear, and can be modeled mathematically as

$$y = Kf + z, \quad (6)$$

where K is a linear operator and z is Gaussian noise [3]. Since the operator K is typically not invertible (i.e., K^{-1} is unbounded), some “regularization” is needed to invert the problem. Unfortunately, traditional regularization methods (e.g., Tikhonov regularization or truncated Singular Value Decomposition [63, 64, 85]) have the undesirable effect that important features to be recovered are lost, as evident in imaging applications where the regularized reconstructions are blurred versions of the original. This phenomenon is of particular concern since, in many situations, the most relevant information to be recovered is indeed contained in edges or other sharp transitions. To address this issue, a number of different methods have been proposed, including hidden Markov models, Total Variation regularization and Anisotropic Diffusion [38, 72, 83, 87]. However, while these methods produce visually appealing results, their rationale is essentially heuristic and they offer no sound theoretical framework to assess the ultimate method performance.

As will be described in the following, recent ideas from sparse representations can be applied to develop a theoretical and computational framework for the regularized inversion of a large class of inverse problems. Specifically, in contrast with more traditional regularization techniques and the other methods mentioned above, the shearlet representation provides a rigorous theoretical framework which is very effective at dealing with a large class of inverse problems and which is especially

effective in the recovery of information associated with edges and other singularities.

3.1 Inverting the Radon Transform

By taking advantage of the ability of the shearlet representation to represent some important classes of operators, a novel method for the regularized inversion of the Radon transform was introduced in [17, 29]. This is a problem of great interest since the Radon transform is the underlying mathematical framework of Computerized Tomography (CT), which has become an essential tool in medical diagnostics and preventive medicine². The Radon transform maps a Lebesgue integrable function f on \mathbb{R}^2 into the set of its line integrals

$$Rf(\theta, t) = \int_{\ell(\theta, t)} f(x) dx,$$

where $\ell(\theta, t)$, with $t \in \mathbb{R}$, $\theta \in S^1$, are the lines $\{x \in \mathbb{R}^2 : x \cdot \theta = t\}$.

Since the shearlet-based approach to invert the Radon transform adapts a number of ideas from the Wavelet–Vaguelette Decomposition (WVD) introduced by Donoho in [21], let us start by briefly recalling the main ideas of the WVD.

Suppose that the operator K in (6) maps the space $L^2(\mathbb{R}^2)$ into the Hilbert space Y . The WVD consists in selecting a well localized orthonormal wavelet basis $\{\psi_{j,m}\}$ of $L^2(\mathbb{R}^2)$ and an appropriate orthonormal basis $U_{j,k}$ of Y so that any $f \in L^2(\mathbb{R}^2)$ can be expressed as

$$f = \sum_{j,k} c_{j,k} [Kf, U_{j,k}] \psi_{j,m}, \quad (7)$$

where $c_{j,k}$ are known scalars and $[\cdot, \cdot]$ is the inner product in Y (we refer to [21] for more detail). It follows that f can be recovered from the observed data Kf and, consequently, an estimate of f can be obtained from the decomposition (7) applied to the noisy data $Kf + z$ using a wavelet thresholding algorithm like those described in Section 2. The main advantage of this approach is that, unlike the classical Singular Value Decomposition (SVD), the basis functions employed in the decomposition formula (7) do not derive entirely from the operator K , but can be chosen to capture most efficiently the features of the object f to be recovered. Indeed this approach turns out to outperform SVD and other standard methods. For functions f in a certain class (specifically, if f is in a certain family of Besov spaces), then the WVD method converges to f with the optimal rate, provided that the thresholding parameters are properly selected [21, 54].

By adapting the WVD approach within the shearlet framework, it is shown in [17, 29] that a function $f \in L^2(\mathbb{R}^2)$ is recovered from the Radon data Rf using the expansion

² More than 72 millions of medical CT scans were performed in the USA, in 2007

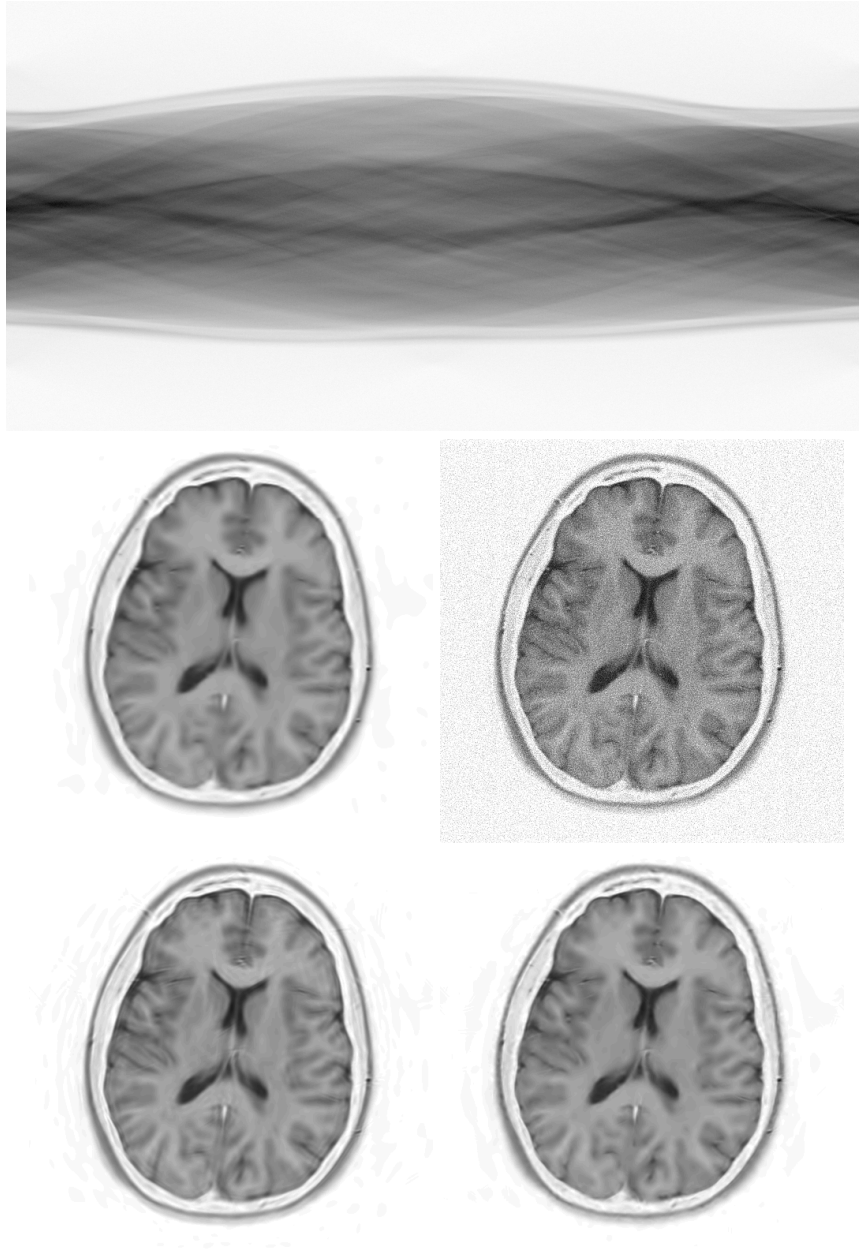


Fig. 5 From the top, clockwise: noisy Radon projections (SNR=34.11 dB); unfiltered reconstruction (SNR=11.19 dB); shearlet-based estimate (SNR=21.68 dB); curvelet-based estimate (SNR=21.26 dB); wavelet-based estimate (SNR=20.47 dB).

$$f = \sum_{j,\ell,m} 2^j [Rf, U_{j,\ell,m}] \psi_{j,\ell,m}, \quad (8)$$

where $\{\psi_{j,\ell,m}\}$ is a Parseval frame of shearlets of $L^2(\mathbb{R}^2)$, and $\{U_{j,\ell,m}\}$ is a related system which is obtained by applying an operator closely related to the Radon transform to the shearlet system. The advantage of this representation is that the shearlet system $\{\psi_{j,\ell,m}\}$ is optimally suited to represent images f containing edges. Thus, if f is a cartoon-like image and the observed Radon data Rf are corrupted by additive Gaussian noise, it is proved that the shearlet-based estimator obtained from shearlet thresholding provides an *optimal* Mean Square Error for the recovery of f . In particular, this approach outperforms the standard WVD as well as other traditional methods, in which cases the MSE has a slower decay rate. With respect to a somewhat similar result based on the curvelet approach [6], the shearlet-based method provides a simpler and more flexible mathematical construction which leads to an improved numerical implementation and performance. A typical application of the shearlet-based regularized reconstruction algorithm is reported in Figure 5 where the method is compared against the curvelet-based algorithm and the wavelet-based one (corresponding to WVD). We refer to [17, 29] for additional numerical tests and details on the algorithm.

A related method of using shearlets to control noise amplification when inverting the Radon transform was presented in [1] in the case when the sampled data is compressed.

3.2 Deconvolution

When image degradations include the blurring introduced by camera motion as well as the noise introduced by the electronics of the system, the model of the degradation can be given as a convolution operation. The process of undoing this convolution operation is commonly known as deconvolution and is known to be an ill-posed inverse problem. To regularize the ill-posed problem, the sparse representation properties of shearlets can be utilized.

The idea of using a sparse representation to regularize deconvolution as well as other inverse problems has been suggested before (see for example [21] and [6]). However, unique to this shearlet approach is the ability for a multi-scale and anisotropic regularization inversion to be done before noise shrinkage [69]. An additional benefit is the use of a cross-validation function to adaptively select the thresholding values.

A digitally recorded image is a finite discrete data set, so an image deconvolution problem can be formulated as a matrix inversion problem. Without loss of generality, we assume the recorded images/arrays are of size $N \times N$. Let γ denote an $N \times N$ array of samples representing a zero mean additive white Gaussian noise with variance σ^2 . Let y denote the observed image and x is assumed to represent the image to be estimated. Then the deconvolution problem can be formulated as

$$\mathbf{y} = H\mathbf{f} + \boldsymbol{\gamma},$$

where \mathbf{y} , \mathbf{f} , and $\boldsymbol{\gamma}$ are $N^2 \times 1$ column vectors representing the arrays y , f , and γ lexicographically ordered, and H is the $N^2 \times N^2$ matrix that models the blur operator. When the assumption of periodic boundary is made, the problem can be described as

$$y(n_1, n_2) = (f * h)(n_1, n_2) + \gamma(n_1, n_2), \quad (9)$$

where $0 \leq n_1, n_2 \leq N - 1$, $*$ denotes circular convolution, and h denotes the point spread function (PSF). In the discrete Fourier transform (DFT) domain, equation (9) reduces to

$$\widehat{y}(k_1, k_2) = \widehat{h}(k_1, k_2)\widehat{f}(k_1, k_2) + \widehat{\gamma}(k_1, k_2), \quad (10)$$

where $\widehat{y}(k_1, k_2)$, $\widehat{h}(k_1, k_2)$, $\widehat{f}(k_1, k_2)$ and $\widehat{\gamma}(k_1, k_2)$ are the discrete Fourier transforms of y , h , f , and γ , respectively, for $-N/2 \leq k_1, k_2 \leq N/2 - 1$. In this formulation, it is evident that if there exist indices (k_1, k_2) where $|\widehat{h}(k_1, k_2)|$ contains values at or near zero, then the system will be ill-conditioned.

Using the regularized inverse operator

$$H'_\alpha(k_1, k_2) = \frac{\widehat{h}(k_1, k_2)}{|\widehat{h}(k_1, k_2)|^2 + \alpha}$$

for some regularizing parameter $\alpha \in \mathbb{R}^+$, an estimate in the DFT domain is given by

$$\widehat{f}_\alpha(k_1, k_2) = \widehat{y}(k_1, k_2)H'_\alpha(k_1, k_2),$$

for $N/2 \leq k_1, k_2 \leq N/2 - 1$. Applying the multi-channel implementation of shearlets, we can adaptively control the regularization parameter to be the best suited for each frequency supported trapezoidal region. Let $g_{j,\ell}$ denote the shearlet filter that will correspond to a given scale j and direction ℓ . The shearlet coefficients of an estimate for a given regularization parameter α can be computed in the DFT domain as

$$\widehat{c}(f_\alpha)_{j,\ell}^S(k_1, k_2) = \widehat{y}(k_1, k_2)\widehat{g}_{j,\ell}(k_1, k_2)H'_\alpha(k_1, k_2),$$

for $N/2 \leq k_1, k_2 \leq N/2 - 1$.

The regularization parameters $\{\alpha\}$ will act to suppress a noise amplification, yet it is desirable to allow a noise amplification as long as the remaining noise level can be adequately controlled by shearlet shrinkage methods.

Shearlet threshold values can be adaptively found by using a generalize cross validation (GCV) as follows: Let \mathbf{y} , \mathbf{f} , and $\boldsymbol{\gamma}$ denote the observed noisy image, original image, and the colored noise so that $\mathbf{y} = \mathbf{f} + \boldsymbol{\gamma}$. The noise is assumed to be second order stationary (i.e. the mean is constant and the correlation between two points depends only on the distance between them).

The soft thresholding function $T_\tau(c)$ is defined to be equal to $c - \tau \text{sign}(c)$ if $|c| > \tau$ and zero otherwise for a given threshold parameter τ . Assuming the noise process $\boldsymbol{\gamma}$ is stationary and that $\langle \boldsymbol{\gamma}, \boldsymbol{\psi}_{j,\ell,m} \rangle$ represents a shearlet coefficient of a random vector $\boldsymbol{\gamma}$ at scale j , direction ℓ , and location m , the following lemma is obtained.

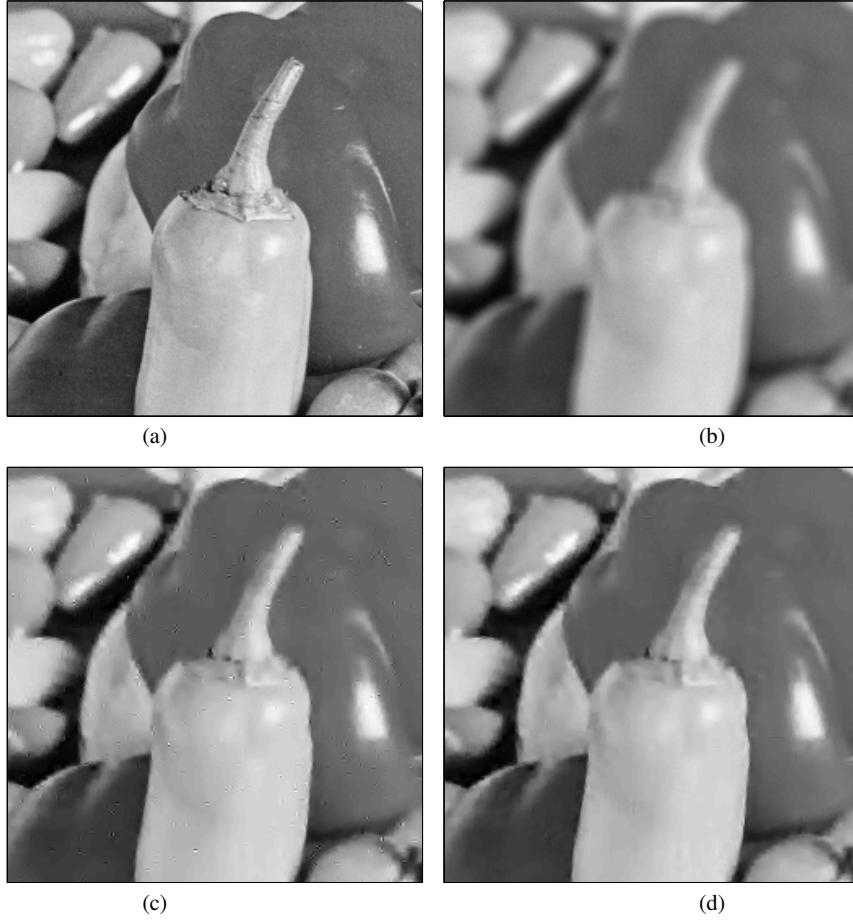


Fig. 6 Details of the image deconvolution experiment with a *Peppers* image. (a) Original image. (b) Noisy blurred image, BSNR=30 dB. (c) ForWaRD estimate, ISNR=4.29 dB. (d) shearlet-based estimate, ISNR=5.42 dB.

Lemma 1 ([69]). $E [|\langle \gamma, \psi_{j,\ell,m} \rangle|^2]$ depends only on the scale j and direction ℓ .

This means the shearlet transform of stationary correlated noise is stationary within each scale and directional component. Let $\mathbf{y}_{j,\ell}$ denote the vector of shearlet coefficients of \mathbf{y} at scale j and direction ℓ . $L_{j,\ell}$ is the number of shearlet coefficients on scale j and direction ℓ , and L is the total number of shearlet coefficients. Given

$$R_{j,\ell}(\boldsymbol{\tau}_{j,\ell}) = \frac{1}{L_{j,\ell}} \|\mathbf{T}_{\boldsymbol{\tau}_{j,\ell}}(\mathbf{y}_{j,\ell}) - \mathbf{f}_{j,\ell}\|^2, \quad (11)$$

the total risk is

$$R(\tau) = \sum_j \sum_\ell \frac{L_{j,\ell}}{L} R_{j,\ell}(\tau_{j,\ell}). \quad (12)$$

This means the minimizing the mean squared error or risk function R can be achieved by minimizing $R_{j,\ell}(\tau_{j,\ell})$ for all j and ℓ . Assuming $L_{j,\ell,0}$ is the total number of shearlet coefficients that were replaced by zero after threshold, we now have the following:

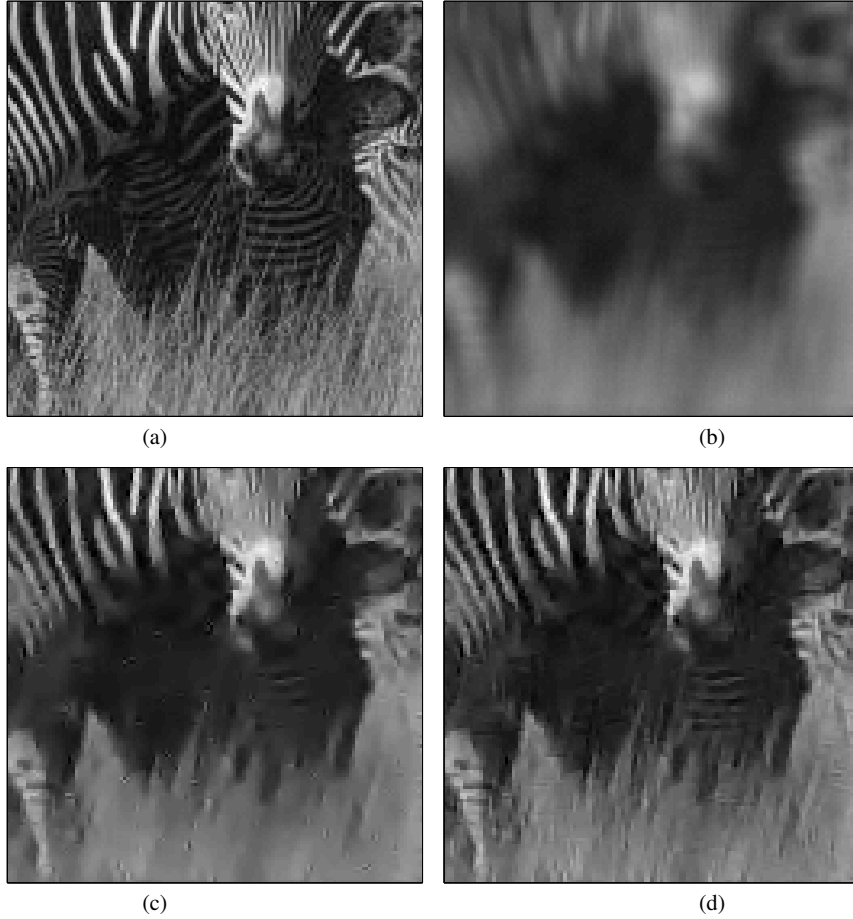


Fig. 7 Details of the image deconvolution experiment with a *Zebra* image. (a) Original image. (b) Noisy blurred image, BSNR=30 dB. (c) ForWaRD estimate, ISNR=5.53 dB. (d) shearlet-based estimate, ISNR=6.03 dB.

Theorem 1 ([69]). *The minimizer of*

$$GCV_{j,\ell}(\tau_{j,\ell}) = \frac{\frac{1}{L_{j,\ell}} \| T_{\tau}(\mathbf{y}_{j,\ell}) - \mathbf{y}_{j,\ell} \|^2}{\left[\frac{L_{j,\ell,0}}{L_{j,\ell}} \right]^2}, \quad (13)$$

is asymptotically optimal for the minimum risk threshold $R_{j,\ell}(\tau_{j,\ell})$ for scale j and directional component ℓ .

This means finding the values $\tau_{j,\ell}$ that minimize the cross validation function $GCV_{j,\ell}$ for each j and ℓ , leads to a shearlet-based estimate that will likely be close to the ideal noise-free image.

Define $\widehat{c}(y)_{j,\ell}^S(k_1, k_2)$ as $\widehat{y}(k_1, k_2)\widehat{g}_{j,\ell}(k_1, k_2)$ for $-N/2 \leq k_1, k_2 \leq N/2 - 1$, and $\widetilde{c}(f_{\alpha})_{j,\ell}^S$ to be the estimate of $c(f_{\alpha})_{j,\ell}^S$ after thresholding the coefficients by using the GCV formula given in (13). That is, for a given α , set $\widetilde{c}(f_{\alpha})_{j,\ell}^S = T_{\tau_{j,\ell}}(c(f_{\alpha})_{j,\ell}^S)$ for $\tau_{j,\ell} = \arg \min_{\tau_{j,\ell}} GCV_{j,\ell}(\tau_{j,\ell})$. Then one option for finding the optimal α for each thresholded set of shearlet coefficients $\widetilde{c}(f_{\alpha})_{j,\ell}^S$ can be found by minimizing the cost function

$$\sum_{k_1} \sum_{k_2} \frac{|\widehat{h}(k_1, k_2)|}{|\widehat{h}(k_1, k_2)|^2 + \eta} \left| \widehat{h}(k_1, k_1) \widetilde{c}(f_{\alpha})_{j,\ell}^S(k_1, k_2) - \widehat{c}(y)_{j,\ell}^S(k_1, k_2) \right|^2,$$

where $\eta = N^2 \sigma^2 / \|c(y)_{j,\ell}^S - \mu(c(y)_{j,\ell}^S)\|_2^2$, $\mu(y)$ denotes the mean of y , and σ is the estimated standard deviation of the noisy blurred image. For other variances and options on finding α , we refer to [69].

We use the improvement in signal-to-noise-ratio (ISNR) to measure the success of the routines and the blurred signal-to-noise ratio (BSNR) to give in understanding of the problem setup. The ISNR is given as

$$ISNR = 10 \log_{10} \left(\frac{\|f - y\|_2^2}{\|f - \widetilde{f}\|_2^2} \right),$$

and the BSNR is given as

$$BSNR = 10 \log_{10} \left(\frac{\|(f * h) - \mu(f * h)\|_2^2}{N^2 \sigma^2} \right),$$

for an $N \times N$ image. Notice that both are measured in decibels.

Applications of the shearlet-based deconvolution algorithm are shown in Figures 6 and 7, where it is compared against the highly competitive wavelet-based deconvolution algorithm known as *Fourier-Wavelet Regularized Deconvolution* (ForWaRD) [65]. Figures 6 and 7 display close-ups of some experiments results where the blur was a 9×9 boxcar blur [65].

3.3 Inverse-Halftoning

Halftoning is a process of rendering an image into a binary (black-and-white) image. Halftoning techniques include error diffusion methods such as those by Floyd-Steinburg and Jarvis *et al.* [86, 47, 36, 48]. At times these halftoned images may need resizing, enhancement, or removal of aliasing artifacts. There may also be a need for these images to be restored to their original gray-scale images for other reasons such as compression or for digital achieving of old newspapers and articles.

Given the Floyd-Steinberg filter

$$h_{FS} = \frac{1}{16} \begin{bmatrix} 0 & \bullet & 7 \\ 3 & 5 & 1 \end{bmatrix},$$

or the Jarvis error filter

$$h_J = \frac{1}{48} \begin{bmatrix} 0 & 0 & \bullet & 7 & 5 \\ 3 & 5 & 7 & 5 & 3 \\ 1 & 3 & 5 & 3 & 1 \end{bmatrix},$$

the quantization error at \bullet is diffused over a causal neighbourhood according to the matrix values. Specifically, each pixel is identified in a raster-scan indexing scheme and the pixel's gray-scale value is made into a binary number by thresholding (1, if the value is greater than or equal to $1/2$, and 0 otherwise). The quantization error is then diffused on neighbouring pixels using the weights from h_{FS} or h_J . Let p and q denote the impulse responses determined by the error diffusion model. The relation between the original $N \times N$ gray-scale image f and the resultant halftone y image can be approximately modeled as:

$$y(n_1, n_2) = (p * f)(n_1, n_2) + (q * v)(n_1, n_2) \quad (14)$$

where $0 \leq n_1, n_2 \leq N - 1$ and v is considered an additive white Gaussian noise even though the process does not involve randomness. In the DFT domain, equation (14) can be written as

$$\hat{y}(k_1, k_2) = \hat{p}(k_1, k_2)\hat{f}(k_1, k_2) + \hat{q}(k_1, k_2)\hat{v}(k_1, k_2),$$

for $-N/2 \leq k_1, k_2 \leq N/2 - 1$. Assuming \hat{h} denotes the DFT of the diffusion filters h_{FS} or h_J , the transfer functions \hat{p} and \hat{q} are given by

$$\hat{p}(k_1, k_2) = \frac{C}{1 + (C - 1)\hat{h}(k_1, k_2)}$$

and

$$\hat{q}(k_1, k_2) = \frac{1 - \hat{h}(k_1, k_2)}{1 + (C - 1)\hat{h}(k_1, k_2)}.$$

where the constant $C = 2.03$ when $h = h_{FS}$ or $C = 4.45$ when $h = h_J$ [48].

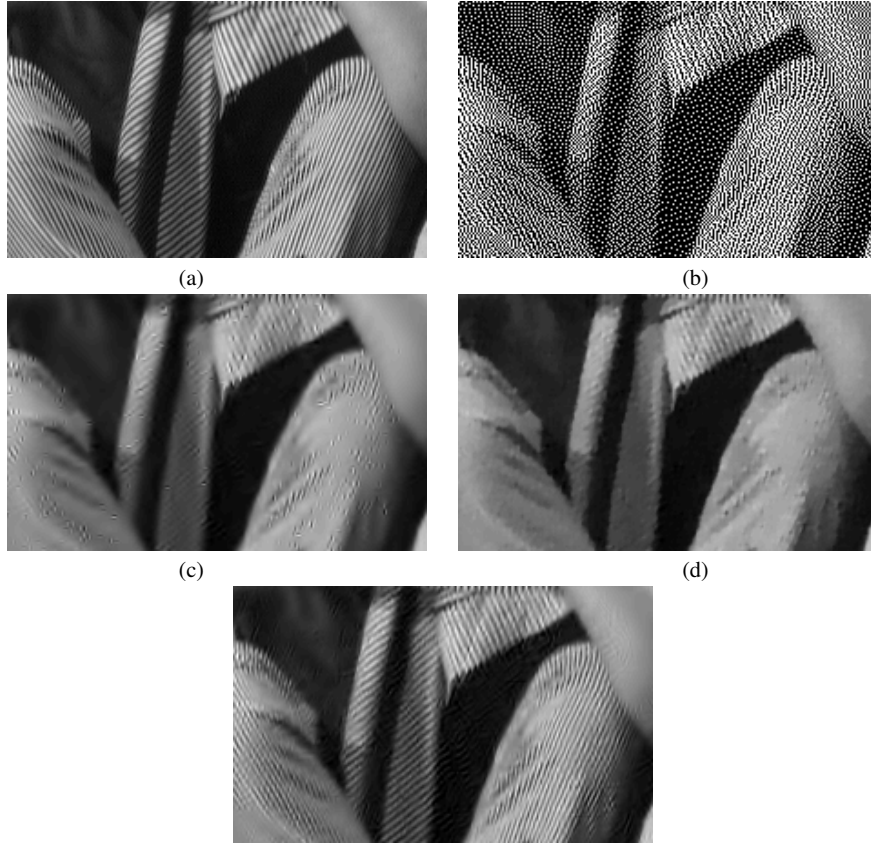


Fig. 8 Close-ups of inverse halftoning experiment with the *Barbara* image. (a) Original image. (b) Floyd-Steinberg halftone. (c) Wavelet-based estimate (SNR = 19.30 dB). (d) LPA-ICI-based estimate (SNR = 19.25 dB). (e) Shearlet-based estimate (SNR = 22.02 dB).

To approximately invert the halftoning process, we use the regularized inverse operator

$$P'_\alpha(k_1, k_2) = \frac{\widetilde{p}(k_1, k_2)}{|\widehat{p}(k_1, k_2)|^2 + \alpha^2 |\widehat{q}(k_1, k_2)|^2} \quad (15)$$

for some regularizing parameter $\alpha \in R^+$. This gives an image estimate in the Fourier domain as

$$\widehat{f}_\alpha(k_1, k_2) = \widehat{y}(k_1, k_2) P'_\alpha(k_1, k_2),$$

for $-N/2 \leq k_1, k_2 \leq N/2 - 1$.

This regularization process can be separated into a shearlet domain as done previously. Assuming $g_{j,\ell}$ denotes the shearlet filter for scale j and direction ℓ , the shearlet coefficients of an estimate of the image for a given regularization parameter

α can be computed in the DFT domain as

$$\widehat{c}(f_\alpha)_{j,\ell}^S(k_1, k_2) = \widehat{y}(k_1, k_2) \widehat{g}_{j,\ell}(k_1, k_2) P'_\alpha(k_1, k_2)$$

for $-N/2 \leq k_1, k_2 \leq N/2 - 1$.

The remaining aspect of this problem is transformed into a form of a denoising problem which can be dealt with by thresholding the estimated shearlet coefficients using the GCV formulation determined previously.

We illustrate the performance of the shearlet inverse halftoning algorithm (see [34] for more details) by using the *Barbara* image. The image was halftoned using the Floyd-Steinberg algorithm and comparisons were done with a wavelet-based method as well as a LPA-ICI-based method [66],[37]. Close-ups of the results are shown in Figure 8.

4 Image Enhancement

Image enhancement is a term describing an improvement of the visual properties of an image for the purpose of interpretation or perception either for human or computer vision systems. Mathematically, given an image whose pixel values are described as the array $y(k_1, k_1)$, it produces an altered image $y_e(k_1, k_2)$ by an application of an enhancement transformation En . That is $y_e(k_1, k_2) = En(y(k_1, k_2))$ for $-N/2 \leq k_1, k_2 \leq N/2 - 1$. If we assume that y is a grayscale image whose pixels are integers ranging from 0 to 255, then a fairly simple enhancement map is $En(t) = 255 - t$, which was used in Fig. 1 to improve the contrast in the images of the shearlet coefficients. Other simple enhancement transforms are based on logarithm, exponential, or piece-wise linear functions. Another powerful and widely used enhancement technique is *histogram equalization* which produces a transformation En such that the histogram of the its pixel values of the enhanced image y_e is evenly distributed. More recently, several image enhancement techniques have been introduced which are based on ideas from multiscale analysis [52, 53, 56, 79].

Image enhancement is frequently used in medical imaging, where it can be helpful to emphasize visual features which are important for medical diagnostic. For example, the enhancement of mammography images can be very useful to improve the visibility of small tumors for early detection. To this goal, many image processing techniques based on multiscale analysis have been found effective, such as the approach of Laine *et al.* [52], which investigates mammography feature analysis using the dyadic wavelet transform, the approach by Strickland *et al.* [81], which uses the undecimated wavelet transform for detecting and segmenting calcifications in mammograms, and the enhancement method of Chang *et al.* [11], which is based on overcomplete multiscale representations.

In this section, we adapt some of the ideas proposed in the literature for image enhancement to construct an image enhancement algorithm based on the shearlet representation. The intuitive idea behind this approach is that, since shearlet co-

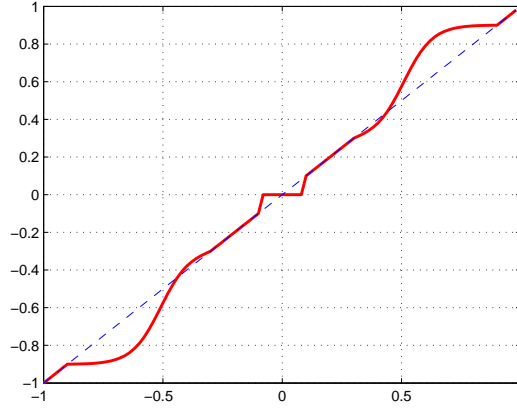


Fig. 9 Enhancement map: $b = 0.35$, $\beta = 10$, $T_1 = 0.1$, $T_2 = 0.3$, $T_3 = 0.9$

efficients are closely related to edges and other essential geometric features, it is possible to enhance such features by controlling the magnitude of the shearlet coefficients. Thus, we introduce an appropriate adaptive nonlinear mapping function on the shearlet coefficients with the goal to amplify weak edges, while keeping the other strong edges intact. We define this nonlinear operator as follows, using the notation $\text{sigm}(t) = (1 + e^{-t})^{-1}$:

$$\begin{aligned}
 En(t) &= 0 \quad \text{if } |t| < T_1 \\
 En(t) &= \text{sign}(t)T_2 + \bar{a}(\text{sigm}(\beta(\bar{t} - b)) - \text{sigm}(-\beta(\bar{t} + b))) \\
 &\quad \text{if } T_2 \leq |y| \leq T_3 \\
 En(t) &= t \quad \text{otherwise}
 \end{aligned} \tag{16}$$

where $t \in [-1, 1]$, $\bar{a} = a(T_3 - T_2)$, $\bar{t} = \text{sign}(t) \frac{|t| - T_2}{T_3 - T_2}$, $0 \leq T_1 \leq T_2 < T_3 \leq 1$, $b \in (0, 1)$, and a , dependent on the gain factor β and b , is defined as

$$a = \frac{1}{\text{sigm}(\beta(1 - b)) - \text{sigm}(-\beta(1 + b))}.$$

In this formulation, the parameters T_1 , T_2 , and T_3 are selected threshold values, and b and β control the threshold and rate of enhancement, respectively. The interval $[T_2, T_3]$ serves as a sliding window for feature selectivity. It can be adjusted to emphasize important features within a specific range. These parameters can be adaptively selected by using the standard deviation of the pixel values for each scale j and direction ℓ . Using this nonlinear function, the shearlet coefficients are point-wise modified for image enhancement by

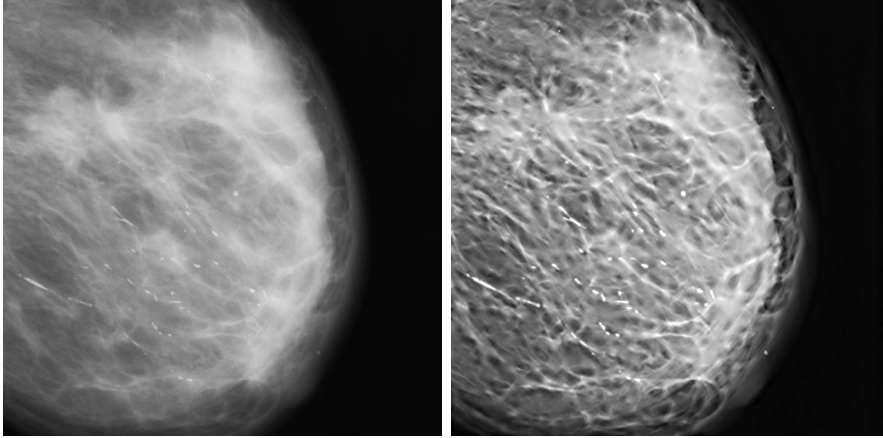


Fig. 10 Image enhanced of mammogram. From left to right: Original mammogram with spiculated masses and calcifications, Image enhanced mammogram using shearlets.

$$\langle y_e, \Psi_{j,\ell,m} \rangle = \max_m (|\langle y, \Psi_{j,\ell,m} \rangle|) \text{En} \left(\frac{\langle y, \Psi_{j,\ell,m} \rangle}{\max_m (|\langle y, \Psi_{j,\ell,m} \rangle|)} \right),$$

where $\max_m (|\langle y, \Psi_{j,\ell,m} \rangle|)$ is the maximum absolute amplitude of $|\langle y, \Psi_{j,\ell,m} \rangle|$ as a function of position m and $\langle y_e, \Psi_{j,\ell,m} \rangle$ denotes the shearlet coefficients of the enhanced image y_e . The resultant enhanced image y_e is found by simply inverting the transform. Figure 9 shows an enhancement map curve representing the enhanced coefficients versus the original coefficients.

Results of the shearlet-based image enhancement method were presented in [68], where they have been applied to enhanced mammogram images. Examples of the application of this algorithm are shown in Fig. 10 and 11, where the enhanced mammograms obtained using shearlets are compared with those obtained using the nonsubsampling wavelet transform (NSWT) and a standard histogram equalization method. In the experiment Fig. 11, we created mathematical models of phantoms to validate our enhancement methods against any false alarms arising from our enhancement techniques. This phantom is a good model for features such as microcalcifications, masses, and spicular objects which occur in real data.

In our experiments, we used 1, 8, 8, 16, 16 directions in the scales from coarser to finer, respectively, as done in [33] for the shearlet decomposition. The standard deviation of pixel values were used to adaptively select the values for T_1 , T_2 , and T_3 . We choose $b = 0.23, 0.14, 0.10, 0.10$ and $\beta = 20, 35, 45, 35$ for the directions in the scales from coarser to finer, respectively. In the first experiment, we enhanced a mammogram image using shearlets as shown in Figure 10. In the second experiment, we blended a normal mammogram with the phantoms and compared our enhancement method with that of the histogram equalization and the NSWT as shown in Figure 11. As in Figure 11, enhancement by shearlets provided a significant improvement in contrast for each feature included in the blended mammogram;

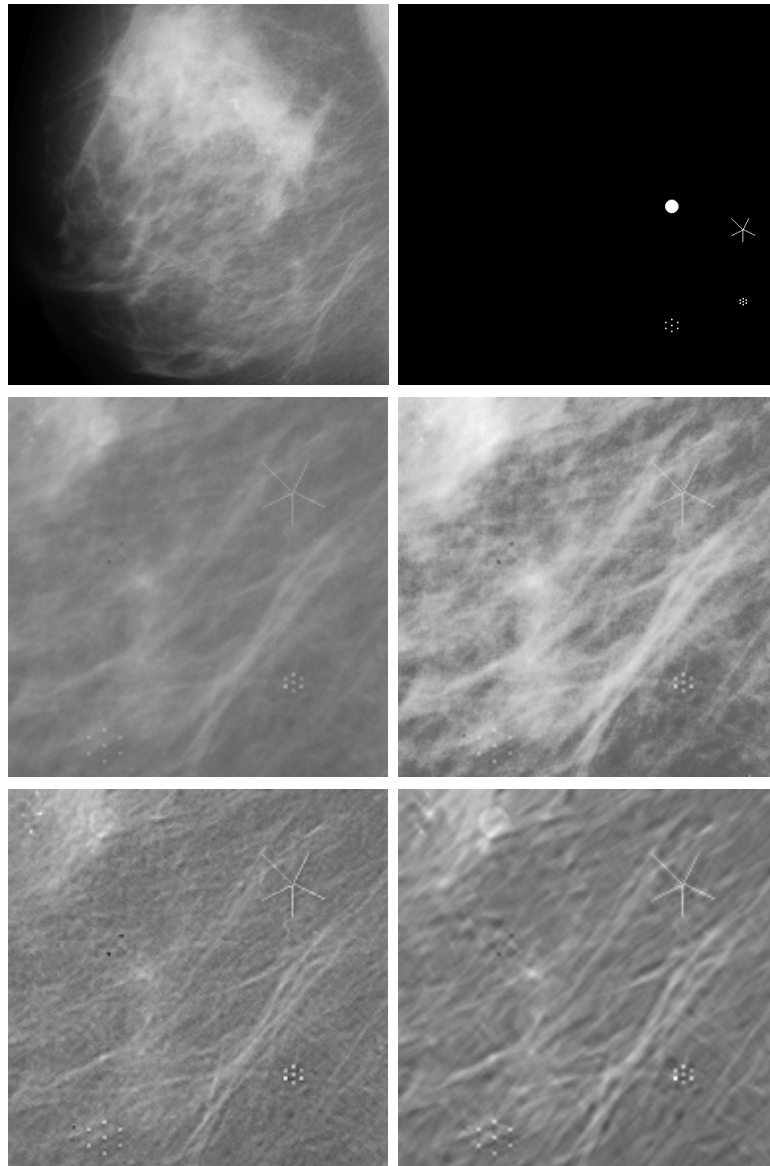


Fig. 11 Enhanced of mammogram with a mathematical phantom inserted. (a) Original mammogram. (b) Mathematical phantom. (c) Region of interest (ROI) image. (d) Enhanced ROI image using histogram equalization. (e) Enhanced ROI image using NSWT. (f) Enhanced ROI image using shearlets.

whereas features such as mass (white disc) is hard to see in the enhanced ROI images obtained by the NSWT and histogram equalization methods.

5 Edge Analysis and Detection

The detection and analysis of edges is a primary task in a variety of image processing and computer vision applications. In fact, since edges are usually the most prominent features in natural images and scientific data, the localization of edges is a fundamental low level task for higher level applications such as shape recognition, 3D reconstruction, data enhancement and restoration.

Edges can be formally characterized as those points of a function u , defined on a domain $\Omega \subset \mathbb{R}^2$, for which the gradient is noticeably large, that is,

$$\{x \in \Omega \subset \mathbb{R}^2 : |\nabla u(x)| > p\},$$

where p is some suitable chosen threshold. It is clear that this simple characterization of edges does not translate directly into an effective edge detection scheme, since images are usually affected by noise and the differential operator is extremely sensitive to noise. As a consequence, in the most common edge detector schemes, to watch out for the interference of noise, the image is first smoothed out or mollified. For example, in the classical Canny edge detection algorithm [8] the image is first convolved with a scalable Gaussian function as

$$u_a = u * G_a,$$

where $G_a(x) = G(a^{-1}x)$, $a > 0$, $x \in \mathbb{R}^2$ and $G(x) = \frac{1}{\pi} e^{-x^2}$. Next, the edge points are identified as the local maxima of the gradient of u_a . Notice that this approach involves a scaling parameter a : as a decreases, the detection of the edge location becomes more accurate; however, as a decreases, also the detector's sensitivity to noise increases. As a result, the performance of the edge detector depends heavily on the scaling factor a (as well as the threshold).

There is an interesting and useful relationship between edge detection and wavelet analysis which was first observed by Mallat, Hwang and Zhong in [60, 61] and can be summarized as follows. Given an image $u \in L^2(\mathbb{R}^2)$, a simple computation shows that its continuous wavelet transform with respect to an admissible real and even function ψ can be written as

$$W_\psi u(a, x) = \int_{\mathbb{R}^2} u(y) D_a \psi(y - x) dy = u * D_a \psi(x).$$

where $D_a \psi(x) = a^{-1} \psi(a^{-1}x)$. In particular, if $\psi = \nabla G$, then

$$\nabla u_a(x) = u * \nabla G_a(x) = u * D_a \psi(x) = W_\psi u(a, x). \quad (17)$$

This shows that the maxima of the magnitude of the gradient of the smoothed image u_a correspond precisely to the maxima of the magnitude of the wavelet transform $W_\psi u(a, x)$. This observation provides a natural mathematical framework for the multiscale analysis of edges which was successfully developed in [60, 61].

5.1 Edge Analysis using Shearlets

The main limitation of the Canny edge detector or the wavelet method described above is that both methods are essentially isotropic and, as a result, are not very efficient at dealing with the anisotropic nature of the edges. The difficulty in accurately identifying the location of edges is particularly evident in the presence of noise and when several edges are close together or cross each other, such as the situation of 2-dimensional projections of 3-dimensional objects [94]. In such cases, the following limitations of traditional edge detectors is particularly evident:

- *Difficulty in distinguishing close edges.* The isotropic Gaussian filtering causes edges running close together to be blurred into a single curve.
- *Poor angular accuracy.* In the presence of sharp changes in curvature or crossing curves, the isotropic Gaussian filtering leads to an inaccurate detection of the edge orientation. This affects the detection of corners and junctions.

To better deal with the edge information, a number of methods were introduced which replace the scalable collection of isotropic Gaussian filters $G_a, a > 0$, in (17) with a family of steerable and scalable anisotropic Gaussian filters such as

$$G_{a_1, a_2, \theta}(x_1, x_2) = a_1^{-1/2} a_2^{-1/2} R_\theta G(a_1^{-1} x_1, a_2^{-1} x_2),$$

where $a_1, a_2 > 0$ and R_θ is the matrix of rotation by the angle θ (see [70, 87, 39]). Unfortunately, the design and implementation of such filters is computationally involved and there is no theoretical setting to decide how to design such family of filters to best capture the edges.

The shearlet framework has the advantage of providing a well justified mathematical setting for efficiently representing the edge information. In fact, as discussed in Chapters 1 and 3 of this volume, the continuous shearlet transform can be applied to precisely characterize the geometric information associated with the edges through its asymptotic behaviour at fine scales. The results can be summarized as follows.

Let an image u be modeled as piecewise smooth function in $\Omega = [0, 1]^2$. That is, we assume that u is smooth everywhere on Ω , except for a collection of finitely many piecewise smooth curves, denoted by Γ , where jump discontinuities may occur. Then the asymptotic decay properties of the continuous shearlet transform \mathcal{SH} of u are as follows [44]:

- If $p \notin \Gamma$, then $|\mathcal{SH}_\psi u(a, s, p)|$ decays rapidly, as $a \rightarrow 0$, for each $s \in \mathbb{R}$.
- If $p \in \Gamma$ and Γ is smooth near p , then $|\mathcal{SH}_\psi u(a, s, p)|$ decays rapidly, as $a \rightarrow 0$, for each $s \in \mathbb{R}$ unless $s = s_0$ is the normal orientation to Γ at p . In this last case, $|\mathcal{SH}_\psi u(a, s_0, p)| \sim a^{\frac{3}{4}}$, as $a \rightarrow 0$.
- If p is a corner point of Γ and $s = s_0, s = s_1$ are the normal orientations to Γ at p , then $|\mathcal{SH}_\psi u(a, s_0, p)|, |\mathcal{SH}_\psi u(a, s_1, p)| \sim a^{\frac{3}{4}}$, as $a \rightarrow 0$. For all other orientations, the asymptotic decay of $|\mathcal{SH}_\psi u(a, s, p)|$ is faster (even if not necessarily “rapid”).

Here by “rapid decay”, we mean that, given any $N \in \mathbb{N}$, there is a $C_N > 0$ such that $|\mathcal{SH}_\psi u(a, s, p)| \leq Ca^N$, as $a \rightarrow 0$. It is also useful to observe that spike-type singularities produce a very different behaviour than jump discontinuities on the decay of the continuous shearlet transform. Consider, for example, a Dirac delta distribution centered at t_0 . In this case a simple calculation (see [49]) shows that

$$|\mathcal{SH}_\psi \delta_{t_0}(a, s, t_0)| \asymp a^{-3/4}, \text{ as } a \rightarrow 0,$$

that is, the continuous shearlet transform of δ_{t_0} , at $t = t_0$ increases at fine scales. The decay is rapid for $t \neq t_0$.

These observations show that the continuous shearlet transform precisely describes the geometric information of the edges and the other singular points of an image. This is in contrast with the wavelet transform which cannot provide any information about the edge orientation.

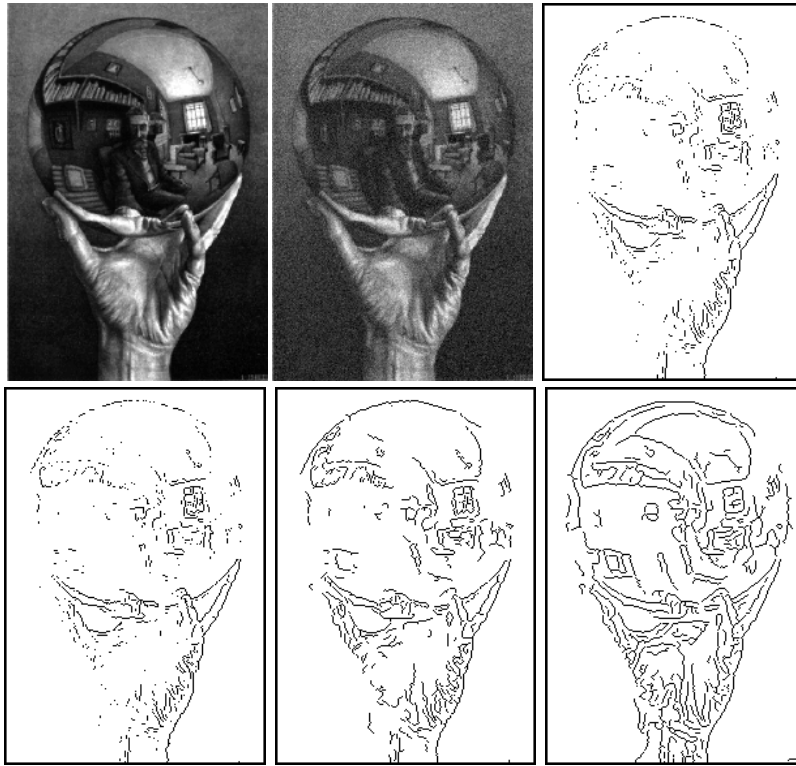


Fig. 12 Results of edge detection methods. From top left, clockwise: Original image, noisy image (PSNR= 25.94 dB), Prewitt result (FOM=0.31), shearlet result (FOM=0.94), wavelet result (FOM=0.59), and Sobel result (FOM=0.32).

5.2 Edge Detection using Shearlets

An algorithm for edge detection based on shearlets was introduced in [90, 91], where a discrete shearlet transform was described with properties specifically designed for this task. In fact, the discrete shearlet transform which was presented above for image denoising, produces large sidelobes around prominent edges³ which interfere with the detection of the edge location. By contrast, the special discrete shearlet transform introduced in [90, 91] is not affected by this issue since the analysis filters are chosen to be consistent with the theoretical results in [44, 45], which require that the shearlet generating function ψ satisfies certain specific symmetry properties in the Fourier domain (this is also discussed in Chapter 3 of this volume).

The first step of the *shearlet edge detector algorithm* consists in selecting the edge point candidates of a digital image $u[m_1, m_2]$. They are identified as those points (\bar{m}_1, \bar{m}_2) which, at fine scales j , are local maxima of the function

$$M_j u[m_1, m_2]^2 = \sum_{\ell} (\mathcal{S}H u[j, \ell, m_1, m_2])^2.$$

Here $\mathcal{S}H u[j, \ell, m_1, m_2]$ denotes the discrete shearlet transform. According to the properties of the continuous shearlet transform summarized above, we expect that, if (\bar{m}_1, \bar{m}_2) is an edge point, the discrete shearlet transform of u will behave as

$$|\mathcal{S}H u[j, \ell, \bar{m}_1, \bar{m}_2]| \sim C 2^{-\beta j},$$

where $\beta \geq 0$. If, however, $\beta < 0$ (in which case the size of $|\mathcal{S}H u|$ increases at finer scales), then (\bar{m}_1, \bar{m}_2) will be recognized as a spike singularity and the point will be classified as noise. Using this procedure, edge point candidates for each of the oriented components are found by identifying the points for which $\beta \geq 0$. Next, a *non-maximal suppression routine* is applied to these points to trace along the edge in the edge direction and suppress any pixel value that is not considered to be an edge. Using this routine, at each edge point candidate, the magnitude of the shearlet transform is compared with the values of its neighbours along the gradient direction (this is obtained from the orientation map of the shearlet decomposition). If the magnitude is smaller, the point is discarded; if it is the largest, it is kept.

Extensive numerical experiments have shown that the shearlet edge detector is very competitive against other classical or state-of-the-art edge detectors, and its performance is very robust in the presence of noise. An example is displayed in Figure 12, where the shearlet edge detector is compared against the wavelet edge detector (which is essentially equivalent to the Canny edge detector) and the Sobel and Prewitt edge detectors. Notice that both the Sobel and Prewitt filters are 2D discrete approximations of the gradient operator. The performance of the edge detectors is assessed using the Pratt's Figure of Merit, which is a fidelity function ranging from 0 to 1, where 1 is a perfect edge detector. This is defined as

³ The same problem occurs if one uses a standard discrete wavelet or curvelet transform

$$\text{FOM} = \frac{1}{\max(N_e, N_d)} \sum_{k=1}^{N_d} \frac{1}{1 + \alpha d(k)^2},$$

where N_e is the number of actual edge points, N_d is the number of detected edge points, $d(k)$ is the distance from the k -th actual edge point to the detected edge point and α is a scaling constant typically set to $1/9$. The numerical test reported in the figures show that the shearlet edge detector consistently yields the best value for FOM.

5.3 Edge Analysis using Shearlets

As observed above, the continuous shearlet transform has the ability to precisely characterize the geometry of the edges. These properties lead directly to a very effective algorithm for the estimation of the edge orientation, which was originally introduced in [91]. Specifically, by taking advantage of the parameter associated with the orientation variable in the shearlet transform, the edge orientations of an image u , can be estimated by searching for the value of the shearing variable s which maximizes $\mathcal{SH}_\psi u(a, s, p)$ at an edge point p , when a is sufficiently small. Discretely, this is obtained by fixing a sufficiently fine scales (i.e., $a = 2^{-2j}$ sufficiently “small”) and computing the index $\tilde{\ell}$ which maximizes the magnitude of the discrete shearlet transform $\mathcal{SH}u[j, \ell, m]$ as

$$\tilde{\ell}(j, m) = \arg \max_{\ell} |\mathcal{SH}u[j, \ell, m]|. \quad (18)$$

Once this is found, the corresponding angle of orientation $\theta_{\tilde{\ell}}(j, m)$ associated with the index $\tilde{\ell}(j, m)$ can be easily computed. As illustrated in [91], this approach leads to a very accurate and robust estimation for the local orientation of the edge curves.

Indeed, the sensitivity of the shearlet transform to the edge orientation is useful for the extraction of landmarks, another imaging application, which is important in problems of classification and retrieval. To illustrate the general principle, consider the simple image in Figure 13 consisting of large smooth regions separated by piecewise smooth curves. The junction point A , where three edges intersect, is certainly the most prominent object in the image, and this can be easily identified by looking at values of the shearlet transform. In fact, if one examines the discrete shearlet transform $\mathcal{SH}u[j_0, l, m_0]$, at a fixed (fine) scale j_0 and locations m_0 , as a function of the shearing parameter l , the plot immediately identifies the local geometric properties of the image. Specifically, as illustrated in Figure 13(b), one can recognize the following four classes of points inside the image. At the junction point $k_0 = A$, the function $|\mathcal{SH}u[j_0, l, m_0]|$ exhibits three peaks corresponding to the orientations of the three edge segments converging into A ; at the point $m_0 = B$, located on a smooth edge, $|\mathcal{SH}u[j_0, l, m_0]|$ has a single peak; at a point $m_0 = D$, inside a smooth region, $|\mathcal{SH}u[j_0, l, m_0]|$ is essentially flat; finally, at a point $m_0 = C$ “close” to an edge, $|\mathcal{SH}u[j_0, l, m_0]|$ exhibit two peaks, but they are much smaller in amplitude than

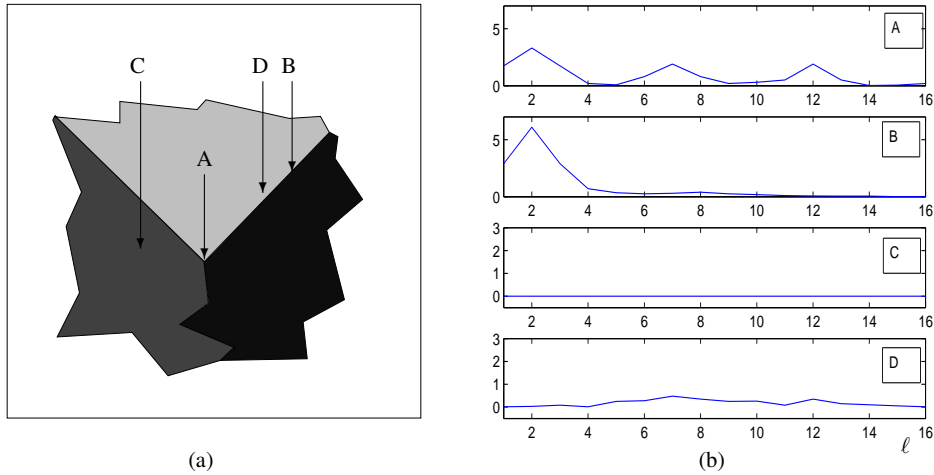


Fig. 13 (a) Test image and representative points A (junction), B (regular edge point), C (smooth region), D (near edge). (b) Magnitude of the Discrete Shearlet Transform, as a function of the orientation parameter ℓ at the locations $m_0 = A, B, C, D$ indicated in (a). Notice the different scaling factor used in the y -axis, for the plots of points C and D .

those for the points A and B . A similar behaviour was observed, as expected, for more general images, even in the presence of noise.

Based on these observations, a simple and effective algorithm for classifying smooth regions, edges, corners and junction points of an image was proposed and validated in [91].

6 Image Separation

Blind source separation is a classical problem in signal processing whose object is the separation of a set of signals from a set of mixed signals, with very little information about the source signals and the mixing process. The traditional techniques for addressing this problem rely on the assumption that the source signals are essentially decorrelated, so that they can be separated into additive subcomponents which are statistically independent. The main weakness of these techniques is that they are very sensitive to noise. On the other hand, recent results have shown that sparse representations such as shearlets can be applied to design extremely robust source separation algorithms [25, 26, 78].

In the following, we describe a very effective algorithm for image separation, recently proposed in [25, 26] which takes advantage of the ability of the shearlet representation in dealing with edge curves and other elongated features. This approach is especially tailored to deal with the situation of images, such as astronomical or

biological images, where it is important to separate pointlike objects from curvelike ones.

6.1 Image model

The class of the images of interest, denoted by J , are modeled as a composition of point- and curve-like objects. That is, a point-like object is a function P which is smooth except for finitely many point singularities and has the form

$$P(x) = \sum_{i=1}^m |x - x_i|^{-\frac{3}{2}}.$$

A curve-like object is a distribution C with delta singularity along a closed curve $\tau : [0, 1] \rightarrow \mathbb{R}^2$. Hence, an image in J will be of the form

$$f = P + C.$$

The goal of the Geometric Separation Problem is to recover P and C from the observed signal f .

The basic idea is to choose a redundant dictionary containing two representations systems Φ_1 , Φ_2 , that is

$$\mathcal{D} = \Phi_1 + \Phi_2,$$

where each system sparsely represents only one of the different components of $f \in J$. Specifically, Φ_1 is chosen to be a Parseval frame of shearlets which, as discussed above, provides optimally sparse approximations of functions which are smooth apart from curve singularities; Φ_2 is chosen to be a smooth wavelet orthonormal basis, which, as known, provides optimally sparse approximations of functions which are smooth apart from point singularities.

6.2 Geometric Separation Algorithm

In the algorithmic approach to the image separation problem devised in [50], an image $f \in J$ is examined at various resolution levels, denoted by f_j , $j \in \mathbb{Z}$, where $f_j = f * F_j$ and F_j is a bandpass filter associated with the frequency band centered at 2^j . Hence, for each $j \in \mathbb{Z}$,

$$f_j = P_j + C_j,$$

where P_j and C_j are the point-like and curve-like components of f_j , respectively, at scale j . At the resolution level j , the following optimization problem is defined:

$$(\widehat{W}_j, \widehat{S}_j) = \operatorname{argmin}_{W_j, S_j} \|\Phi_1^T S_j\|_1 + \|\Phi_2^T W_j\|_1 \quad \text{subject to } f_j = S_j + W_j, \quad (19)$$

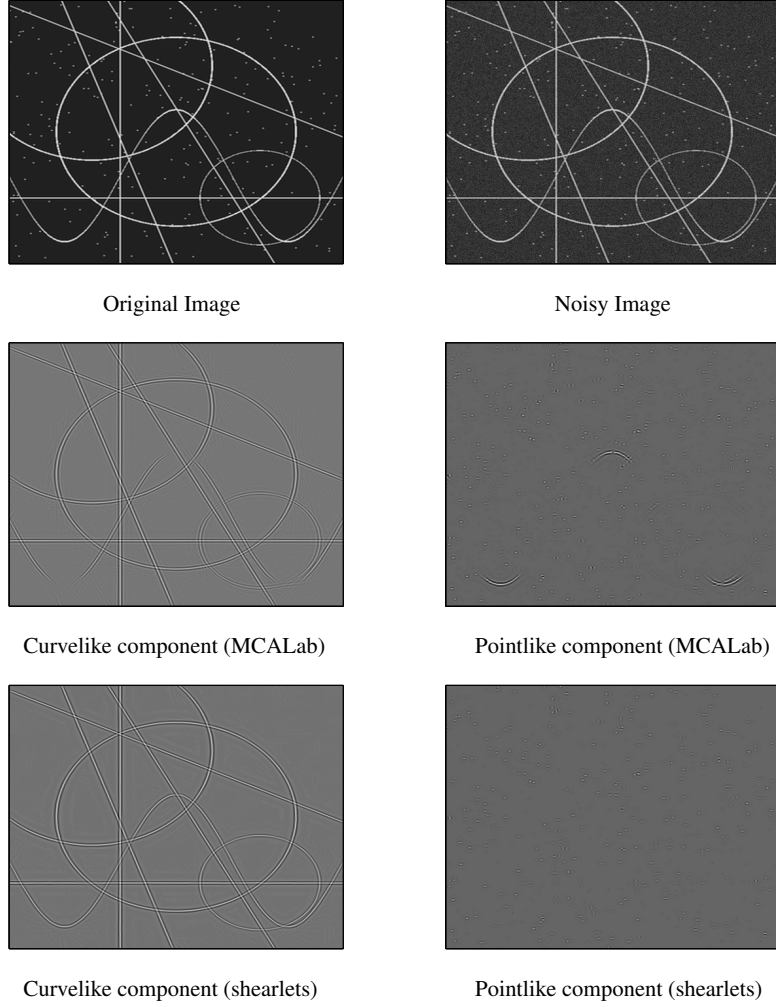


Fig. 14 Example of geometric separation on a noisy synthetic image. The shearlet-based geometric separation algorithm is compared against MCALab.

where $\Phi_1^T S_j$ and $\Phi_2^T W_j$ are the shearlet and wavelet coefficients of the signals S_j and W_j , respectively.

The following theoretical result from [26] ensures the convergence of the geometric separation problem at fine scales.

Theorem 2. Let $(\widehat{W}_j, \widehat{S}_j)$ be the solutions to the optimization problem (20) for each scale j . Then

$$\lim_{j \rightarrow \infty} \frac{\|\widehat{W}_j - P_j\|_2 + \|\widehat{S}_j - C_j\|_2}{\|P_j\|_2 + \|C_j\|_2} = 0.$$

This shows that the components P_j and C_j of f_j are recovered with asymptotically arbitrarily high precision at fine scales.

In practice, an image f is not purely a sum of a point-like and a curve-like components, and contains an additional part that can be modeled as a noise term. In this situation, one can modify the optimization problem (20) as follows

$$(\widehat{W}_j, \widehat{S}_j) = \operatorname{argmin}_{W_j, S_j} \|\Phi_1^T S_j\|_1 + \|\Phi_2^T W_j\|_1 + \lambda \|f_j - W_j - S_j\|_2, \quad (20)$$

subject to $f_j = S_j + W_j$. Notice that this type of expression is sometimes called an *infimal convolution* in the image processing literature. In this modified form, the additional noisy component in the image is characterized by the property that it can not be represented sparsely by either one of the two representation systems and, thus, will be allocated to the residual term $(f_j - W_j - S_j)$.

An example of the application of the shearlet-based geometric separation algorithm to a noisy image is illustrated in Figure 14, where the result is compared to the MCALab algorithm [35], another separation algorithm which employs a combination of curvelet and wavelet representations. The figure shows that the shearlet-based approach is very effective at separating the pointlike and curvelike components of the image, and it produces significantly less artifacts than MCALab. We refer to [26] for a more detailed discussion.

7 Shearlets analysis of 3D Data

A number of results have recently appeared dealing with the application of shearlet-based methods for the analysis and processing of 3D data sets. Similar to the 2D case, the nearly optimally sparse approximation properties of 3D shearlet representations can be exploited for data denoising and feature extraction. As expected, dealing with 3D data sets entails more challenges in terms of memory storage so that particular attention is required to devise numerical efficient implementations.

A 3D Discrete Shearlet Transform (3D DST) was proposed in [51] and tested on video denoising. The algorithm follows essentially the ideas of the 2D discrete shearlet algorithm and can be summarized as follows. First, the data in the frequency domain are divided into three pyramidal regions, each one aligned with one of the orthogonal axis. The directional filtering stage is based on computing the DFT in the pseudopolar domain. In particular, in the first pyramidal region, this is defined as $(u, v, w) = (\xi_1, \frac{\xi_2}{\xi_1}, \frac{\xi_3}{\xi_1})$. Hence, at each fixed resolution level, the 3D DST algorithm proceeds as follows.

- The multiscale filter stage decomposes f_d^{j-1} into a low-pass f_d^j and high-pass \hat{f}_d^j array.
- \hat{f}_d^j is rearranged onto a pseudo-polar grid.
- A directional band pass filtering is applied on the pseudo-polar data.
- The pseudo-polar data is converted back to a Cartesian formulation and the inverse DFT is computed.

The algorithm runs in $O(N^3 \log(N))$ operations.

A shearlet thresholding routine based on the 3D DST algorithm (3DSHEAR) was applied to a problem of video denoising and the performance was compared against the following state-of-the-art algorithms: the Dual Tree Wavelet Transform (DTWT) and Surfacelets (SURF). We also compared against the 2D discrete shearlet transform (2DSHEAR), which was applied frame by frame, to illustrate the benefit of using a 3D transform, rather than a 2D transform acting on each frame. Figure 15 shows a side-by-side comparison of the denoising algorithm performance on a typical frame extracted from the video sequence Mobile. Additional comparisons and discussion can be found in [51].

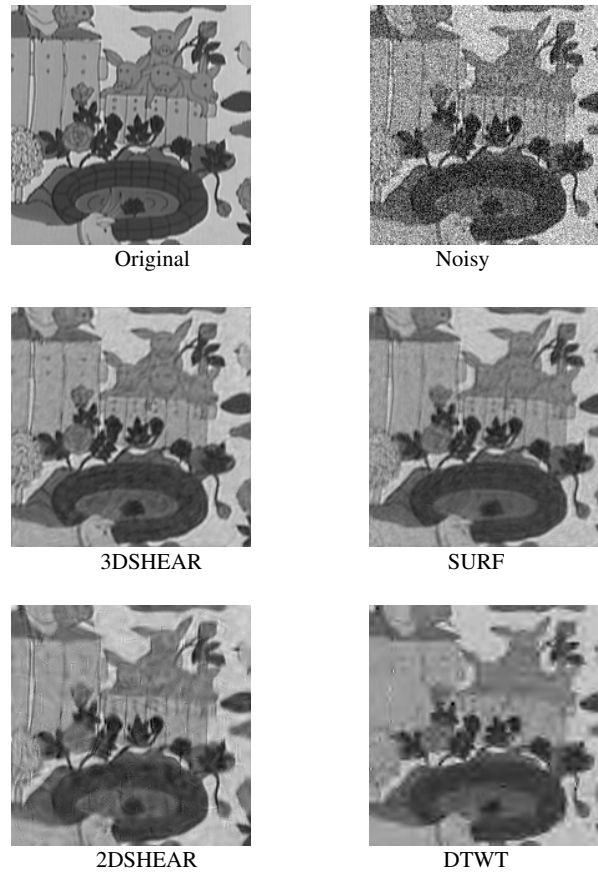


Fig. 15 Side by side comparison of video denoising algorithms, illustrated on a frame extracted from a video sequence. The 3D Discrete Shearlet Transform (3DSHEAR) is compared against the Dual Tree Wavelet Transform (DTWT), the Surfacelet Transform (SURF) and the 2D Discrete Shearlet Transform (2DSHEAR)

A shearlet-based method of analyzing videos from multiple views to autonomously estimate the kinematic state of an object was developed in [75]. In particular, a target's kinematics state was parametrized as a vector

$$\mathbf{X} = [\mathbf{s} \quad \boldsymbol{\tau}]$$

whose respective corresponding elements were spatial position $\mathbf{s} = [x, y, z]$ and rotational orientation $\boldsymbol{\tau} = [h, p, r]$. Information from a Bayesian filter was merged to stabilize 2D recognition and tracking so that observation and object were concurrent. In this application, the shearlet transform was used to extract image features reliably. In particular, it was used to determine two dimensional locations in the midst of illumination changes and discontinuities.

In an effort to improve the state estimation routine, a continuous-type of 3D shearlet transform was developed to analyze video data. In this case, the 3D shearlet transform was being used for detecting surface boundaries [76]. An illustration of the power of using a 3D shearlet surface/edge detector routine over a slice by slice detection of the 2D shearlet edge detector is given in Figure 7. In this example, a solid spherical harmonic of order 2 and degree 7 is generated with a gradient shading applied to each slice. The same slices are analyzed by the 2D shearlet edge detector for comparisons. The images show the contour surface plot of this spherical harmonic and image slices through the center aligned with the x,y, and z axis.

8 Additional Applications

Among the other areas of image processing that benefit from the use of the shearlet representation, we also recall image fusion and inpainting. In image fusion, the goal is to process and synthesize information provided by various sensors. A novel image fusion algorithm based on shearlets and local energy was recently proposed in [57], where it was shown that this approach outperforms traditional methods by preserving more details in the source images and further improving the subjective quality of fused image. In [19], another shearlet-based image fusion method was developed for panchromatic high resolution images and multispectral images, and it was shown that it provides superior performance in terms of spatial resolution and preservation of spectral information. For the applications of inpainting, which can be described as an interpolation or estimation problem with missing data, a shearlet-based technique was recently presented in [41].

Acknowledgements

D.L. acknowledges support from NSF grants DMS 1008900 and DMS (Career) 1005799.

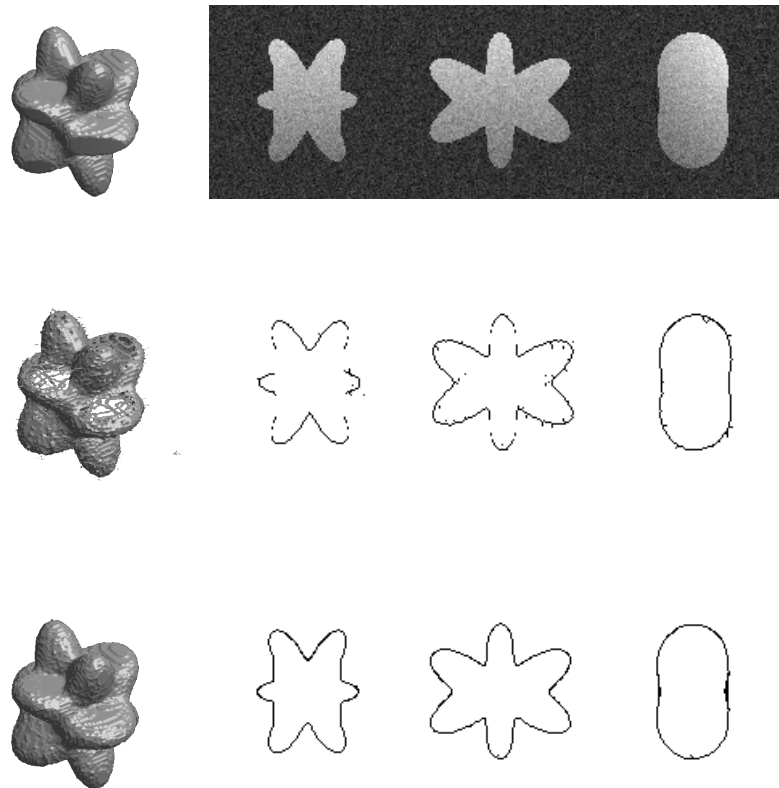


Fig. 16 Illustrations of 3D edge detection of a solid spherical harmonic of order 2 and degree 7 with a 2D gradient shading applied. The images on the left display the 3D contour surface plots and the images on the left display the slices of sphere through the center.

References

1. J. Aelterman, H. Q. Luong, B. Goossens, A. Pizurica, W. Philips, *Compass: a joint framework for Parallel Imaging and Compressive Sensing in MRI*, Image Processing (ICIP), 17th IEEE International Conference on (2010), 1653–1656.
2. A. Averbuch, R. R. Coifman, D. L. Donoho, M. Israeli, and Y. Shkolnisky, *A framework for discrete integral transformations I - the pseudo-polar Fourier transform*, SIAM Journal on Scientific Computing **30**(2) (2008), 764–784.
3. M. Bertero, *Linear inverse and ill-posed problems*, Advances in Electronics and Electron Physics (P.W. Hawkes, ed.), Academic Press, New York, 1989.
4. L. Blanc-Feraud, P. Charbonnier, G. Aubert, and M. Barlaud, *Nonlinear image processing: modelling and fast algorithm for regularization with edge detection*, Proc. IEEE ICIP-95, **1** (1995), 474–477.

5. P. J. Burt, E. H. Adelson, *The Laplacian pyramid as a compact image code*, IEEE Trans. Commun. **31** (4) (1983), 532–540.
6. E. J. Candès, and D. L. Donoho, *Recovering edges in ill-posed inverse problems: optimality of curvelet frames*, Annals Stat. **30**(3) (2002), 784–842.
7. E. J. Candès and F. Guo, *New multiscale transforms, minimum total variation synthesis: applications to edge-preserving image reconstruction*, Signal Proc. **82**(11) (2002), 1519–1543.
8. F. J. Canny, *A computational approach to edge detection*, IEEE Trans. Pattern Anal. Machine Intell. **8**(6) (1986), 679–698.
9. H. Cao, W. Tian, C. Deng, *Shearlet-based image denoising using bivariate model*, Progress in Informatics and Computing (PIC), 2010 IEEE International Conference on **2** (2010), 818–821.
10. T. Chan, J. Shen, *Image Processing And Analysis: Variational, PDE, Wavelet, And Stochastic Methods*, SIAM, Philadelphia (2005).
11. C. Chang, A. F. Laine, *Coherence of Multiscale Features for Contrast Enhancement of Digital Mammograms*, IEEE Trans. Info. Tech. in Biomedicine **3**(1) (1999), 32–46.
12. G. Chang, B. Yu and M. Vetterli, *Adaptive Wavelet Thresholding for Image Denoising and Compression*, IEEE Trans. Image Processing, **9** (2000), 1532–1546.
13. X. Chen, C. Deng, S. Wang, *Shearlet-Based Adaptive Shrinkage Threshold for Image Denoising*, E-Business and E-Government (ICEE), 2010 International Conference on (2010), 1616–1619.
14. X. Chen, H. Sun, C. Deng, *Image Denoising Algorithm Using Adaptive Shrinkage Threshold Based on Shearlet Transform*, Frontier of Computer Science and Technology, 2009, Fourth International Conference on (2009), 254–257.
15. R. R. Coifman and A. Sowa, *Combining the calculus of variations and wavelets for image enhancement*, Appl. Comput. Harmon. Anal., **9** (2000), 1–18.
16. F. Colonna, G. R. Easley, *Generalized discrete Radon transforms and their use in the ridgelet transform*, Journal of Mathematical Imaging and Vision, **23** (2005), 145–165.
17. F. Colonna, G. R. Easley, K. Guo, and D. Labate, *Radon Transform Inversion using the Shearlet Representation*, Appl. Comput. Harmon. Anal. **29**(2) (2010), 232–250.
18. C. Deng, H. Sun, X. Chen, *Shearlet-Based Adaptive Bayesian Estimator for Image Denoising*, Frontier of Computer Science and Technology, 2009, Fourth International Conference on (2009), 248–253.
19. C. Deng, S. Wang, X. Chen, *Remote Sensing Images Fusion Algorithm Based on Shearlet Transform*, Environmental Science and Information Application Technology, 2009, International Conference on **3** (2009), 451–454.
20. D. L. Donoho, *Unconditional bases are optimal bases for data compression and for statistical estimation*, Appl. Comput. Harmon. Anal. **1**(1) (1993), 100–115.
21. D. L. Donoho, *Nonlinear solution of linear inverse problems by wavelet-vaguelette decomposition*, Appl. Comput. Harmon. Anal. **2** (1995), 101–126.
22. D. L. Donoho, *De-noising by soft thresholding*, IEEE Trans. Info. Theory **41** (1995), 613–627.
23. D. L. Donoho and I. M. Johnstone, *Ideal spatial adaptation via wavelet shrinkage*, Biometrika **81** (1994), 425–455.
24. D. L. Donoho and I. M. Johnstone, *Adapting to unknown smoothness via wavelet shrinkage*, J. Amer. Stat. Assoc. **90**(432) (1995), 1200–1224.
25. D. L. Donoho and G. Kutyniok, *Geometric Separation using a Wavelet-Shearlet Dictionary*, SampTA-09 (Marseille, France, 2009), Proc., 2009.
26. D. L. Donoho and G. Kutyniok, *Microlocal analysis of the geometric separation problem*, preprint.
27. D. L. Donoho, M. Vetterli, R. A. DeVore, and I. Daubechies, *Data compression and harmonic analysis*, IEEE Trans. Inform. Theory, **44** (1998), 2435–2476.
28. S. Durand and J. Froment, *Reconstruction of wavelet coefficients using total variation minimization*, SIAM J. Sci. Comput., **24**(5) (2003), 1754–1767.

29. G. R. Easley, F. Colonna, and D. Labate, *Improved Radon Based Imaging using the Shearlet Transform*, Proc. SPIE, Independent Component Analyses, Wavelets, Unsupervised Smart Sensors, Neural Networks, Biosystems, and Nanoengineering VII, **7343**, Orlando, April 2009.
30. G. R. Easley, D. Labate, *Critically sampled composite wavelets*, Signals, Systems and Computers, 2009 Conference Record of the Forty-Third Asilomar Conference on (2009), 447–451.
31. G. R. Easley, D. Labate, and F. Colonna, *Shearlet-Based Total Variation for Denoising*, IEEE Trans. Image Processing, **18**(2) (2009), 260–268.
32. G. R. Easley, D. Labate, and W-Q Lim, *Sparse Directional Image Representations using the Discrete Shearlet Transform*, Appl. Comput. Harmon. Anal. **25**(1) (2008), 25–46.
33. G. R. Easley, V. Patel, D. M. Healy, Jr., *An M-channel Directional Filter Bank Compatible with the Contourlet and Shearlet Frequency Tiling*, Wavelets XII, Proceedings of SPIE, San Diego, CA (2007), 26–30.
34. G. R. Easley, V. M. Patel, and D. M. Healy, Jr., *Inverse halftoning using a shearlet representation*, Proc. of SPIE Wavelets XIII, **7446**, San Diego, August 2009.
35. M. J. Fadilli, J. L. Starck, M. Elad, and D. L. Donoho, *MCALab: reproducible research in signal and image decomposition and inpainting*, IEEE Comput. Sci. Eng. Mag., 2009, to appear.
36. R. W. Floyd and L. Steinberg, *An adaptive algorithm for spatial grayscale*, Proc. Soc. Image Display **17**(2) (1976), 75–77.
37. A. Foi, V. Katkovnik, K. Egiazarian, and J. Astola, *Inverse halftoning based on the anisotropic LPA-ICI deconvolution*, Proc. Int. TICSP Workshop Spectral Methods Multirate Signal Processing, (Vienna, Austria) (2004), 49–56.
38. D. Geman and C. Yang, *Nonlinear image recovery with half-quadratic regularization*, IEEE Trans. Image Proc. **4** (1995), 932–946.
39. J. Geusebroek, A. W. M. Smeulders, and J. van de Weijer, *Fast anisotropic Gauss filtering*, IEEE Trans. Image Proc. **8** (2003), 938–943.
40. G. Gilboa, Y. Y. Zeevi, and N. Sochen, *Texture preserving variational denoising using an adaptive fidelity term*, Proc. VLISM, Nice (2003), 137–144.
41. R. Gomathi and A. Kumar, *An efficient GEM model for image inpainting using a new directional sparse representation: Discrete Shearlet Transform*, Computational Intelligence and Computing Research (ICCIC), 2010 IEEE International Conference on (2010), 1–4.
42. B. Goossens, J. Aelterman, H. Luong, A. Pizurica, and W. Philips, *Efficient design of a low redundant Discrete Shearlet Transform*, Local and Non-Local Approximation in Image Processing, 2009, International Workshop on (2009), 112–124.
43. K. Guo and D. Labate, *Optimally sparse multidimensional representation using shearlets*, SIAM J. Math. Anal. **39** (2007), 298–318.
44. K. Guo and D. Labate, *Characterization and analysis of edges using the continuous shearlet transform*, SIAM J. Imaging Sciences **2** (2009), 959–986.
45. K. Guo, D. Labate and W. Lim, *Edge analysis and identification using the continuous shearlet transform*, Appl. Comput. Harmon. Anal. **27** (2009), 24–46.
46. Q. Guo, S. Yu, X. Chen, C. Liu, and W. Wei, *Shearlet-based image denoising using bivariate shrinkage with intra-band and opposite orientation dependencies*, Computational Sciences and Optimization, 2009, International Joint Conference on, **1** (2009), 863–866.
47. J. Jarvis, C. Judice, and W. Ninke, *A survey of techniques for the display of continuous tone pictures on bilevel displays*, Comput. Graph and Image Proc. **5** (1976), 13–40.
48. T. D. Kite, B. L. Evans, and A. C. Bovik, *Modeling and quality assessment of halftoning by error diffusion*, IEEE Trans. Image Proc. **9** (2000), 909–922.
49. G. Kutyniok and D. Labate, *Resolution of the wavefront set using continuous shearlets*, Trans. Amer. Math. Soc. **361** (2009), 2719–2754.
50. G. Kutyniok and W. Lim, *Image separation using shearlets*, preprint.
51. D. Labate and P. Negi, *3D Discrete shearlet transform and video denoising*, Proc. SPIE, 2011.

52. A. F. Laine, S. Schuler, J. Fan, and W. Huda, *Mammographic feature enhancement by multi-scale analysis*, IEEE Trans. Med. Imag. **13**(4) (1994), 725–752.
53. A. F. Laine and X. Zong, *A multiscale sub-octave wavelet transform for de-noising and enhancement*, Wavelet Applications, Proc. SPIE, Denver, CO, August 6-9, 1996, **2825**, 238–249.
54. N. Lee and B J Lucier, *Wavelets methods for inverting the Radon transform with noisy data*, IEEE Trans. Image Proc. **10**(1) (2001), 79–94.
55. W. Q. Lim, *The discrete shearlet transform: a new directional transform and compactly supported shearlet frames*, Image Proc. IEEE Transactions on **19**(5) (2010), 1166–1180.
56. J. Lu and D. M. Healy, Jr., *Contrast enhancement via multi-scale gradient transformation*, Wavelet Applications, Proc. SPIE, Orlando, FL, April 5-8, 1994.
57. L. Lü , J. Zhao, and H. Sun, *Multi-focus image fusion based on shearlet and local energy*, Signal Processing Systems (ICSPS), 2010 2nd International Conference on, **1** (2010), V1-632–V1-635.
58. J. Ma and M. Fenn, *Combined complex ridgelet shrinkage and total variation minimization*, SIAM J. Sci. Comput., **28**(3) (2006), 984–1000.
59. S. Mallat, *A Wavelet Tour of Signal Processing*, Academic Press, San Diego, 1998.
60. S. Mallat and W. L. Hwang, *Singularity detection and processing with wavelets*, IEEE Trans. Inf. Theory **38**(2) (1992), 617–643.
61. S. Mallat and S. Zhong, *Characterization of signals from multiscale edges*, IEEE Trans. Pattern Anal. Mach. Intell. **14**(7) (1992), 710–732.
62. Y. Meyer, *Oscillating Patterns in Image Processing and Nonlinear Evolution Equations*, AMS, Providence, 2001.
63. F. Natterer, *The Mathematics of Computerized Tomography*, Wiley, New York, 1986.
64. F. Natterer and F. Wübbeling, *Mathematical Methods in Image Reconstruction*, SIAM Monographs on Mathematical Modeling and Computation, Philadelphia, 2001.
65. R. Neelamani, H. Choi, and R. G. Baraniuk, *ForWaRD: Fourier-wavelet regularized deconvolution for ill-conditioned systems*, IEEE Trans. Image Proc. **52**(2) (2004), 418–433.
66. R. Neelamani, R. Nowak, and R. Baraniuk, *Model-based inverse half-toning with Wavelet-Vaguelette Deconvolution*, Proc. IEEE Int. Conf. Image Proc. (2000), 973–976.
67. V. M. Patel, G. R. Easley, and R. Chellappa, *Multiscale directional filtering of noisy InSAR phase images*, Proc. SPIE, Independent Component Analyses, Wavelets, Neural Networks, Biosystems, and Nanoengineering VII **7703**, Orlando, April 2010.
68. V. M. Patel, G. R. Easley, and D. M. Healy, Jr., *A new multiresolution generalized directional filter bank design and application in image enhancement*, Proc. IEEE International Conference on Image Proc., San Diego, October 2008, 2816–2819.
69. V. M. Patel, G. R. Easley, and D. M. Healy, Jr., *Shearlet-based deconvolution*, IEEE Trans. Image Proc. **18**(12) (2009), 2673–2685.
70. P. Perona, *Steerable-scalable kernels for edge detection and junction analysis*, Image Vis. Comput. **10** (1992), 663–672.
71. P. Perona and J. Malik, *Scale-space and edge detection using anisotropic diffusion*, IEEE Trans. Pattern Anal. Mach. Intel. **12** (1990), 629–639.
72. L. Rudin, S. Oscher, and E. Fatemi, *Nonlinear total variation based noise removal algorithms*, Phys. D **60** (1992), 259–268.
73. O. Scherzer, M. Grasmair, H. Grossauer, M. Haltmeier and F. Lenzen, *Variational Methods in Imaging*, Springer, Applied Mathematical Sciences 167, 2009.
74. O. Scherzer and J. Weickert, *Relations between regularization and diffusion filtering* Journal of Mathematical Imaging and Vision **12**(1) (2000), 43–?63.
75. D. A. Schug and G. R. Easley, *Three dimensional Bayesian state estimation using shearlet edge analysis and detection*, Communications, Control and Signal Processing (ISCCSP), 2010 4th International Symposium on (2010), 1–4.
76. D. A. Schug, G. R. Easley, and D. P. O’Leary, *Three-dimensional shearlet edge analysis*, Proc. SPIE, Independent Component Analyses, Wavelets, Neural Networks, Biosystems, and Nanoengineering IX, Orlando, April 2011.

77. J. L. Starck, E. J. Candès, and D. L. Donoho, *The curvelet transform for image denoising*, IEEE Trans. Im. Proc. **11** (2002), 670–684.
78. J. L. Starck, M. Elad, and D. L. Donoho, *Image decomposition via the combination of sparse representation and a variational approach*, IEEE Trans. Image Proc. **14** (2005), 1570–1582.
79. J. L. Starck, F. Murtagh, E. J. Candès, and D. L. Donoho, *Gray and color image contrast enhancement by the curvelet transform*, IEEE Trans. Imag. Proc. **12**(6) (2003), 706–717.
80. G. Steidl, J. Weickert, T. Brox, P. Mrázek, and M. Welk, *On the equivalence of soft wavelet shrinkage, total variation diffusion, total variation regularization, and SIDes*, SIAM J. Numer. Anal. **42** (2004), 686–713.
81. R. N. Strickland and H. I. Hahn, *Wavelet Transforms for Detecting Microcalcifications in Mammograms*, IEEE Trans. on Med. Imag. **15**(2) (1996), 218–229.
82. H. Sun and J. Zhao, *Shearlet Threshold Denoising Method Based on Two Sub-swarm Exchange Particle Swarm Optimization*, Granular Computing (GrC), 2010 IEEE International Conference on (2010), 449–452.
83. S. Teboul, L. Blanc-Feraud, G. Aubert, and M. Barlaud, *Variational approach for edge-preserving regularization using coupled PDEs*, IEEE Trans. Image Proc. **7** (1998), 387–397.
84. W. Tian, H. Cao, and C. Deng, *Shearlet-based adaptive MMSE estimator for image denoising*, Intelligent Computing and Intelligent Systems (ICIS), 2010 IEEE International Conference on **2** (2010), 689–692.
85. A. N. Tikhonov, *Solution of incorrectly formulated problems and the regularization method*, Soviet Math. Doklady **4** (1963), 1035–1039.
86. R. Ulichney, *Digital Halftoning*, MIT Press, Cambridge, MA, 1987.
87. J. Weickert, *Foundations and applications of nonlinear anisotropic diffusion filtering*, Z. Angew. Math. Mechan. **76** (1996), 283–286.
88. J. Weickert, *Anisotropic Diffusion in Image Processing*, Teubner, Stuttgart, 1998.
89. M. Welk, G. Steidl and J. Weickert, *Locally analytic schemes: A link between diffusion filtering and wavelet shrinkage*, Appl. and Comput. Harmon. Anal. **24** (2008), 195–224.
90. S. Yi, D. Labate, G. R. Easley, and H. Krim, *Edge detection and processing using shearlets*, Proc. IEEE Int. Conference on Image Proc., San Diego, October 12-15, 2008.
91. S. Yi, D. Labate, G. R. Easley, and H. Krim, *A Shearlet approach to edge analysis and detection*, IEEE Trans. Image Proc. **18**(5) (2009), 929–941
92. X. Zhang, X. Sun, L. Jiao, and J. Chen, *A Non-Local Means Filter with Translating Invariant Shearlet Feature Descriptors*, Wireless Communications Networking and Mobile Computing (WiCOM), 2010 6th International Conference on (2010), 1–4.
93. X. Zhang, Q. Zhang, and L. Jiao, *Image Denoising with Non-Local Means in the Shearlet Domain*, Multi-Platform/Multi-Sensor Remote Sensing and Mapping (M2RSM), 2011 International Workshop on (2011), 1–5.
94. D. Ziou and S. Tabbone, *Edge Detection Techniques An Overview*, Internat. J. Pattern Recognition and Image Anal. **8**(4) (1998), 537–559.

Index

- BayesShrink algorithm, 9
- deconvolution, 17
- digital shearlet transform, 5, 7
- discrete shearlet transform, 5
 - 3D case, 36
 - implementation, 7
- edge analysis, 32
- edge detection, 28
 - Canny edge detector, 28
 - shearlet edge detection, 31, 38
 - wavelet edge detection, 28
- geometric separation, 34
- halftoning, 22
- histogram equalization, 24
- image denoising, 2
- image enhancement, 24
- image separation, 33
- pseudo-polar DFT, 6
- Radon transform, 15
- shearlet thresholding, 7, 11, 17, 37
- Total Variation regularization, 10
- wavelet thresholding, 3, 15
 - hard thresholding, 3
 - soft thresholding, 3
- wavelet–vaguelette decomposition (WVD), 15