Shearlet Based Total Variation for Denoising

Glenn R. Easley, Demetrio Labate, and Flavia Colonna

Abstract—We propose a shearlet formulation of the total variation (TV) method for denoising images. Shearlets have been mathematically proven to represent distributed discontinuities such as edges better than traditional wavelets and are a suitable tool for edge characterization. Common approaches in combining wavelet-like representations such as curvelets with TV or diffusion methods aim at reducing Gibbs-type artifacts after obtaining a nearly optimal estimate. We show that it is possible to obtain much better estimates from a shearlet representation by constraining the residual coefficients using a projected adaptive total variation scheme in the shearlet domain. We also analyze the performance of a shearlet-based diffusion method. Numerical examples demonstrate that these schemes are highly effective at denoising complex images and outperform a related method based on the use of the curvelet transform. Furthermore, the shearlet-TV scheme requires far fewer iterations than similar competitors.

Index Terms—Shearlets, curvelets, total variation, diffusion, regularization, denoising.

I. INTRODUCTION

Restoring images contaminated by measurement errors that cause noise is an important problem in signal processing. Common powerful techniques for image denoising are based on wavelets as well as on total variation (TV) and diffusion.

By relying on certain smoothness assumptions, wavelet theory can be used to provide an effective way to denoise image. For example, if the image is assumed to be a function of class $C^2(\mathbb{R}^2)$ away from a C^2 edge (namely, a composite of a C^2 function plus an indicator function of a set whose boundary is C^2), then the nonlinear approximation of f consisting of the N largest wavelet coefficients has error rate $O(N^{-1})$. Thus, a good approximation can be obtained from some of the largest wavelet coefficients and a denoised estimate of the image can be made by removing the wavelet coefficients whose absolute value is below a specified noise level [1]. This approach, however, often leads to the formation of Gibbs-type (or ringing) artifacts around sharp discontinuities, due to the elimination of small wavelet coefficients that should have been retained. In addition, this technique as well as other sophisticated wavelet coefficient reduction schemes (e.g. [2]) do not necessarily remove all high-noise values (outliers). Although new wavelet extensions such as curvelets [3], [4], [5], [6] (which inspired the source of many of these extensions) and shearlets [7], [8] have a better approximation rate, they may also suffer from the same types of effects.

TV and diffusion-based methods are other powerful tools for denoising and greatly reduce these ringing effects. It is generally understood that they have superior denoising performance when applied to simple classes of images with no textures, such as images of conic shapes with flat colors. These methods, however, often produce approximations that are reminiscent of oil-paintings when applied to images that contain complex textures and shading.

To improve upon these methods, combinations of these routines have been proposed (e.g. [9], [10], [11], [12], [13]). The main goal of these methods was to reduce the the Gibbs-type ringing by adding a constraint on the non-retained coefficients. In an opposite approach, wavelet-inspired concepts were used in [14] to improve the performance and computational efficiency of TV-based methods. Other PDE-based methods influenced by concepts from wavelet theory have been developed in [15], [16], and [17].

In this article, we propose a method for denoising images based on combining the new tight frame of shearlets with TV techniques. A key feature is that the discrete shearlet transform has many flexible attributes that lead to better stability and reduced Gibbs-type artifacts. A closely related approach in [19] suggested com-

System Planning Corporation 1000 Wilson Boulevard, Arlington, VA 22209, USA (geasley@sysplan.com)

North Carolina State University, Campus Box 8205, Raleigh, NC 27695, USA (dlabate@scsu.edu)

George Mason University, 4400 University Drive, Fairfax, VA 22030, USA (fcolonna@gmu.edu)

EDICS: 2-REST, 2-WAVP.

bining the tight frame of curvelets with nonlinear anisotropic diffusion. The results given in [19] indicate that this technique is highly effective. We shall demonstrate that our method based on combining shearlets with TV performs better than this curvelet-based technique. Furthermore, the number of iterations is significantly reduced. In some cases, the reduction in the number of iterations is nearly six-fold.

In Section II, we give a brief overview of the TV and the nonlinear diffusion methods. The shearlet transform and its implementation are described in Section III. In Section IV, we present a new method which exploits the best features of shearlets and TV to obtain superior denoising capabilities. In Section V, we discuss the experimental results of the comparison among different state-of-the-art techniques, and show that the method we propose yields significantly better outcomes. The concluding remarks are given in Section VI.

II. Total Variation and Diffusion

Let Ω be a bounded region in $\mathbb{R}^2$. The *total variation* of a function $u \in C^1(\Omega)$ is defined as

$$TV(u) = \int_\Omega \|\nabla u\| \, dA,$$

where $\nabla u = \left(\frac{\partial u}{\partial x_1}, \frac{\partial u}{\partial x_2} \right)$ and $\| \ \|$ is the standard Euclidean norm.

A common TV technique for the purpose of denoising is based on minimizing the functional

$$F(u) = \int_\Omega \|\nabla u\| \, dA + \frac{\lambda}{2} \int_\Omega (u - u_0)^2 \, dA,$$

where u is the estimated image, u_0 is the noisy image, and $\lambda \in \mathbb{R}_+$ is a penalty parameter (see [20]). The associated Euler-Lagrange equation is

$$-\nabla \cdot \left(\frac{\nabla u}{\|\nabla u\|} \right) + \lambda(u - u_0) = 0,$$

with the Neumann boundary condition $\frac{\partial u}{\partial n} = 0$ on $\partial \Omega$. To improve stability, the term $\|\nabla u\|$ is replaced by $\|\nabla u\|_\alpha = \sqrt{\|\nabla u\|^2 + \alpha}$, where α is a positive parameter.

A method for finding the minimizer of the functional F is based on looking for the steady state solution of

$$\frac{\partial u}{\partial t} = \nabla \cdot \left(\frac{\nabla u}{\|\nabla u\|} \right) - \lambda(u - u_0)$$

with the boundary condition $\frac{\partial u}{\partial n} = 0$ on $\partial \Omega$, where t is interpreted as an artificial time-marching parameter [20]. Other techniques for finding the solution to the Euler-Lagrange equation include duality-based methods [21], [22], [23].

The TV method described above is a special case of the method based on minimizing the functional

$$\int_\Omega \phi(\|\nabla u\|) \, dx \, dy + \frac{\lambda}{2} \int_\Omega (u - u_0)^2 \, dx \, dy,$$

where $\phi \in C^2(\mathbb{R})$ is an even regularization function [24]. The solution is obtained by solving

$$\frac{\partial u}{\partial t} = \nabla \cdot \left(\frac{\phi'(\|\nabla u\|)}{\|\nabla u\|} \nabla u \right) - \lambda(u - u_0) \quad (1)$$

subjected to the Neumann boundary condition. For $\lambda = 0$ and $\lim_{x \to \infty} \phi'(x)/x = 0$, equation (1) is a special case of the Perona and Malik diffusion equation

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

where $\rho(x) = \phi'(x)/x$ [25].

In diffusion, the auxiliary function ρ is used to control the amount of smoothing. In regions where the gradient ∇u is small, which may correspond to noise or the lack of an edge, the diffusion process is strong. On the other hand, in regions where ∇u is large, which are likely to correspond to the location of an edge, the diffusion process is weak or non-existent.

In this context, as we shall see, shearlets behave like the gradient. Indeed, shearlets can be used to detect the presence of an edge [26].

III. Shearlet Transform

The *continuous wavelet transform* W_ψ provides a decomposition of a signal over dilated and translated versions of a fixed waveform ψ. More precisely, for a fixed $\psi \in L^2(\mathbb{R}^n)$ ($n \in \mathbb{N}$), this is defined as the mapping W_ψ with domain $L^2(\mathbb{R}^n)$ such that for $f \in L^2(\mathbb{R}^n)$

$$W_\psi f(a, t) = \int_{\mathbb{R}^n} f(x) \, \overline{\psi_{a,t}(x)} \, dx, \qquad (2)$$

where $\psi_{a,t}(x) = a^{-n/2} \psi(a^{-1}(x - t))$, $a > 0$ and $t \in \mathbb{R}^n$. If the function ψ satisfies the admissibility condition

$$\int_0^\infty |\hat{\psi}(a\xi)|^2 \frac{da}{a^n} = 1 \quad \text{for a.e. } \xi \in \mathbb{R}^n,$$

then ψ is called a *wavelet*, and any $f \in L^2(\mathbb{R}^n)$ can be recovered via the reproducing formula:

$$f = \int_0^\infty \int_{\mathbb{R}^n} \langle f, \psi_{a,t} \rangle \, \psi_{a,t} \, dt \, \frac{da}{a^{2n}}.$$

One of the most remarkable properties of the wavelet transform is its ability to identify the singularities of a signal. In fact, if f is a function which is smooth apart from a discontinuity at a point x_0, the transform $W_\psi f(a,t)$ will signal the location of the singularity by its asymptotic decay at fine scales. More precisely, provided ψ is a "nice" wavelet, then $W_\psi f(a,t)$ decays rapidly as $a \to 0$, unless t is near x_0 [27]. This property allows one to resolve the *singular support* of f, that is, to identify the set of points where f is not regular.

However, the continuous wavelet transform is unable to provide additional information about the *geometry* of the set of singularities of f. This is due the fact that this transform is isotropic (the analyzing elements $\psi_{a,t}$ are obtained by applying the same dilation factor for all coordinate directions) and, as a result, it has a very limited ability to resolve edges and other distributed discontinuities which usually occur in multidimensional data.

To deal with multidimensional signals effectively, one has to introduce a transform with a superior directional sensitivity. For example, this can be obtained by employing a non-isotropic version of the continuous wavelet transform (2) called the *continuous shearlet transform*, introduced by the authors and their collaborators in [28] and [29]. In dimension $n = 2$, this is defined as the mapping

$$\mathcal{SH}_\psi f(a,s,t) = \langle f, \psi_{a,s,t} \rangle,$$

where $a > 0, s \in \mathbb{R}, t \in \mathbb{R}^2$, and the analyzing elements $\psi_{a,s,t}$, called *shearlets*, are given by

$$\psi_{a,s,t}(x) = |\det M_{a,s}|^{-\frac{1}{2}} \psi(M_{a,s}^{-1} x - t), \quad (3)$$

where $M_{a,s} = \begin{pmatrix} a & \sqrt{a}s \\ 0 & \sqrt{a} \end{pmatrix}$. Observe that $M_{a,s} = B_s A_a$, where $A_a = \begin{pmatrix} a & 0 \\ 0 & \sqrt{a} \end{pmatrix}$ and $B_s = \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix}$. Hence to each matrix $M_{a,s}$ are associated two distinct actions: an *anisotropic* dilation produced by the matrix A_a, and a *shearing* produced by the non-expansive matrix B_s.
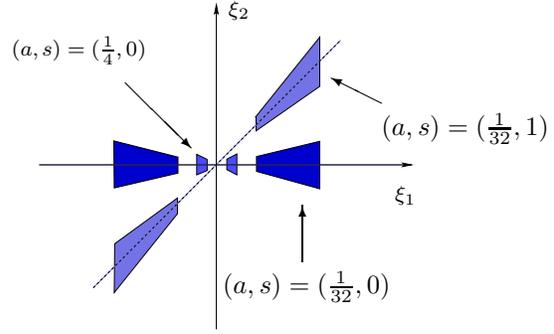

Fig. 1.   Frequency support of the horizontal shearlets (left) and vertical shearlets (right) for different values of a and s.

The generating function ψ is well localized and satisfies appropriate admissibility conditions [28], [29], so that for each $f \in L^2(\mathbb{R}^2)$, we have

$$f = \int_{\mathbb{R}^2} \int_{-\infty}^\infty \int_0^\infty \langle f, \psi_{a,s,t} \rangle \, \psi_{a,s,t} \, \frac{da}{a^3} \, ds \, dt. \quad (4)$$

In particular, for $\xi = (\xi_1, \xi_2) \in \mathbb{R}^2$, $\xi_2 \neq 0$, $\hat{\psi}$ is chosen to be of the form

$$\hat{\psi}(\xi) = \hat{\psi}_1(\xi_1) \, \hat{\psi}_2(\tfrac{\xi_2}{\xi_1}),$$

where $\hat{\psi}_1$, $\hat{\psi}_2$ are smooth functions with supports contained in $[-2, -\frac{1}{2}] \cup [\frac{1}{2}, 2]$ and $[-1, 1]$, respectively. In the frequency domain, the value of $\hat{\psi}_{a,s,t}(\xi_1, \xi_2)$ is given by

$$a^{\frac{3}{4}} \, e^{-2\pi i \xi t} \, \hat{\psi}_1(a\,\xi_1) \, \hat{\psi}_2(a^{-\frac{1}{2}}(\tfrac{\xi_2}{\xi_1} - s)),$$

and, thus, each shearlet $\hat{\psi}_{a,s,t}$ has support in the set

$$\{(\xi_1, \xi_2) : \xi_1 \in [-\tfrac{2}{a}, -\tfrac{1}{2a}] \cup [\tfrac{1}{2a}, \tfrac{2}{a}], \, |\tfrac{\xi_2}{\xi_1} - s| \le \sqrt{a}\}.$$

This shows that each element $\psi_{a,s,t}$ has frequency support on a pair of trapezoids, at various scales, symmetric with respect to the origin and oriented along a line of slope s. The support becomes increasingly thin as $a \to 0$. As a result, the shearlets form a collection of well-localized waveforms at various scales, orientations and locations, controlled by a, s, t, respectively. The frequency supports of some representative shearlets are illustrated in Figure 1. We refer to [28] and [29] for additional details about this construction.
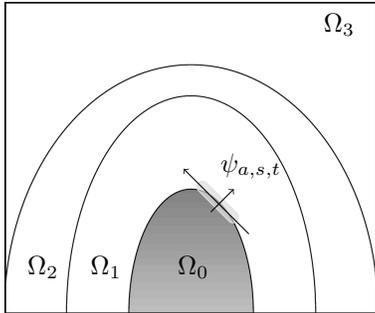
Fig. 2.   An illustration of an image that is $E^{1,3}(\Omega)$ and the essential support of a shearlet $\psi_{a,s,t}$ that intersects an edge.

Thanks to their analytic and geometric properties, the continuous shearlets are able to capture very precisely the geometry of edges. These properties are examined in detail in [29], where it was shown that the asymptotic decay rate of the continuous shearlet transform $\mathcal{SH}_\psi f(a,s,t)$, for $a \to 0$ (fine scales), can be used to signal both the location and the orientation of the edges of an image u. For example, let Ω be a bounded subset of $\mathbb{R}^2$ such that

$$\Omega = \bigcup_{n=1}^{L} \Omega_n \cup \Gamma, \qquad (5)$$

where:
  1. the sets Ω_n, $n = 1, \dots, L$, are pairwise disjoint domains (i.e. connected open sets);
  2. $\Gamma = \bigcup_{n=1}^{L} \partial_\Omega \Omega_n$, where each boundary $\partial_\Omega \Omega_n$ (with respect to the relative topology in Ω) is a smooth C^3 curve of finite length.
Consider the space of images $E^{1,3}(\Omega)$ defined as the collection of functions defined on Ω of the form

$$f(x) = \sum_{n=1}^{L} f_n(x)\, \chi_{\Omega_n}(x) \text{ for } x \in \Omega \backslash \Gamma$$

where $f_n \in C_0^1(\Omega)$, for each $n = 1, \dots, L$, with bounded partial derivatives. This model of images is typically adopted in PDE/variational methods. An illustration is shown in Figure 2.

For each x in a C^3 component of Γ, we define the *jump of f at x*, denoted by $[f]_x$, as

$$[f]_x = \lim_{\epsilon \to 0^+} f(x + \epsilon\, v_x) - f(x - \epsilon\, v_x)$$

where v_x is a unit normal vector along Γ at x. Then we have the following result.

Theorem III.1: Let $f \in E^{1,3}(\Omega)$. Suppose that, for $t \in \Gamma$, in a neighborhood of $t = (t_1, t_2)$ the boundary curve is parametrized as $(E(t_2), t_2)$.
  1. If $s = -E'(t_2)$, then there is a positive constant K such that

$$\lim_{a \to 0} a^{-\frac{3}{4}} \mathcal{SH}_\psi f(a,s,t) = K\, |[f]_t|, \quad \text{as } a \to 0.$$

  2. If $s \neq -E'(t_2)$, or if $t \notin \Gamma$, then

$$\lim_{a \to 0} a^{-\frac{3}{4}} \mathcal{SH}_\psi f(a,s,t) = 0.$$

This shows that the asymptotic decay of the continuous shearlet transform of f is slowest for t on the boundary Γ and s corresponding to the normal orientation to Γ at t. This information can be used to detect the boundary set for a function $f \in E^{1,3}(\Omega)$.

The above property indicates that shearlets can be used to locate and characterize edges. In fact, Theorem III.1 shows that the shearlet coefficients of large magnitude will come from edges. Furthermore, the decay rate across scales can be used to distinguish between noise spikes and edges.

A. Discrete shearlet transform

By sampling the continuous shearlet transform $\mathcal{SH}_\psi f(a,s,t)$ on appropriate discretizations of the scaling, shear, and translation parameters $(a,s,t) \in \mathbb{R}^+ \times \mathbb{R} \times \mathbb{R}^2$, one obtains a discrete transform which is associated to a Parseval (tight) frame for $L^2(\mathbb{R}^2)$ [8]. To do this, we parametrize the anisotropic matrices A_a by dyadic numbers and the shear matrices B_s by integers. Specifically, choosing $a = 2^{-j}$ and $s = -\ell$ with $j, \ell \in \mathbb{Z}$, we obtain the collection of matrices $M_{2^{-j}, -\ell}$. Note that $M_{2^{-j}, -\ell}^{-1} = M_{2^j, \ell}$, where

$$M_{2^j, \ell} = \begin{pmatrix} 2^j & \ell 2^{j/2} \\ 0 & 2^{j/2} \end{pmatrix}$$

$$= \begin{pmatrix} 1 & \ell \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 2^j & 0 \\ 0 & 2^{j/2} \end{pmatrix} = B_0^\ell A_0^j,$$

having denoted by A_0 and B_0 the matrices $\begin{pmatrix} 2 & 0 \\ 0 & \sqrt{2} \end{pmatrix}$, and $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$, respectively. Also, let us replace the continuous translation variable $t \in \mathbb{R}^2$ by a point in the discrete lattice $\mathbb{Z}^2$. From (3) we obtain the discrete system of shearlets $\{\psi_{j,\ell,k}\}$, where

$$\psi_{j,\ell,k} = |\det A_0|^{j/2}\, \psi(B_0^\ell A_0^j x - k),$$

for $j, \ell \in \mathbb{Z}, k \in \mathbb{Z}^2$. For appropriate choices of a well localized generating function ψ, the discrete shearlets form a Parseval frame for $L^2(\mathbb{R}^2)$. That is, for each $f \in L^2(\mathbb{R}^2)$, we have the reproducing formula:

$$f = \sum_{j,\ell \in \mathbb{Z}, k \in \mathbb{Z}^2} \langle f, \psi_{j,\ell,k} \rangle \, \psi_{j,\ell,k}, \qquad (6)$$

with convergence in the L^2 sense. Notice that this equation is the discrete analogue of (4).

The discrete shearlets inherit from their continuous counterpart a special ability to deal with multidimensional functions. In fact, it is proved in [30] that the shearlet are essentially optimal in approximating two-dimensional functions with smooth edges.

Theorem III.2: Let f be defined on a bounded domain $\Omega \subset \mathbb{R}^2$, and suppose that f is C^2 apart from discontinuities along C^2 curves. Let $\tilde{f}_N$ be the approximation of f obtained by taking the N largest coefficient $|\psi_{j,\ell,k}|$ in the shearlets expansion of f given by (6). Then the asymptotic approximation error is given by

$$\|f - \tilde{f}_N\|^2 \le C \, (\log N)^3 \, N^{-2}, \quad N \to \infty.$$

This approximation error is close to optimal (the optimal rate being $O(N^{-2})$) and it significantly outperforms traditional wavelets, whose corresponding asymptotic error decays as N^{-1}. The curvelets, introduced by Candès and Donoho, are the only other system known to satisfy similar approximation properties. Notice, however, that the curvelets, unlike the shearlets, are not an affine system since they are not obtained by applying dilations and translations to a single generator.

The shearlet approach has some similarities with the contourlets [31], [32]. However, the contourlets are a purely discrete construct, while the shearlet theory provides a nice transition from the continuous to the discrete setting, and this is important to derive useful mathematical estimates such as those in Theorems III.1 and III.2. We refer to [33] for additional comments and comparisons.

The reason for the optimal approximation properties of shearlets lies in their special ability to capture distributed discontinuities. Indeed, let f be a two-dimensional function which is smooth apart from some edges. As a closer look at the proof of Theorem III.2 would show, at fine scales, the shearlet coefficients $\langle f, \psi_{j,\ell,k} \rangle$ are negligible unless an edge of f passes near the point $x_{j,\ell,k} = A_0^{-j} B_0^{-\ell} k$ and its orientation corresponds to the value of ℓ. This shows that f can be very well approximated using a shearlet representation $\tilde{f}$ as

$$\sum_{j,\ell,k \in M_1} \langle f, \psi_{j,\ell,k} \rangle \psi_{j,\ell,k} + \sum_{j,\ell,k \in M_2} \langle f, \psi_{j,\ell,k} \rangle \psi_{j,\ell,k}$$

where M_1, M_2 are "small" index sets, the first corresponding to the coarse scale coefficients associated with the smooth regions of f and the second to the fine scale coefficients associated with the edges of f.

These observations suggest that the shearlet decomposition can be used to characterize certain function spaces which are particularly useful in image analysis. In fact, images are usually realized as L^2 objects, while the more noticeable features in images belong to proper subclasses of L^2. Edges, for example, are well quantified within the smaller space BV of functions of bounded variation. The other relevant features, including homogeneous regions, texture and other oscillatory patterns, belong to certain intermediate classes lying between the larger space $L^2(\mathbb{R}^2)$ and the smaller space $BV(\mathbb{R}^2)$. By adapting the ideas proposed in some recent papers such as [17], the shearlet representation can be used to provide an appropriate decomposition of the space $L^2(\mathbb{R}^2)$ into a sequence of spaces associated with a *hierarchy* of scales. Namely, by using a shrinkage approach to remove shearlets below a threshold, $f \approx \sum_{|c_\mu| > \lambda} c_\mu \psi_\mu$, one is able to extract the features of f above a certain scale. This provides a multiscale description of f in an intermediate scale of spaces between BV and L^2. Notice that, unlike traditional wavelets, this approach is able to optimally capture directional features and is thus particularly suitable for image representation.

B. Implementation

The discrete shearlets described above provide a nonuniform angular covering of the frequency plane. For numerical implementation, it is preferred to reformulate the shearlet transform as follows. Define $\hat{\psi}^{(0)}(\xi) = \hat{\psi}_1(\xi_1)\hat{\psi}_2(\frac{\xi_2}{\xi_1})$,
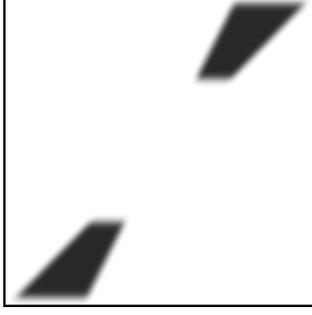
Fig. 3. An image of the shearlet filters $(v_j * w_{j,\ell})$ in the frequency domain.

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

$$A_1 = \begin{pmatrix} \sqrt{2} & 0 \\ 0 & 2 \end{pmatrix}, \quad B_1 = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}.$$

Given $\psi_{j,\ell,k}^{(d)}(x) = 2^{\frac{3j}{2}}\psi^{(d)}(B_d^\ell A_d^j x - k)$, for $d = 0, 1$, the *shearlet transform* is the correspondence mapping $f \in L^2(\mathbb{R}^2)$ into $\langle f, \psi_{j,\ell,k}^{(d)}\rangle$, where $j \geq 0, -2^j \leq \ell \leq 2^j - 1$, and $k \in \mathbb{Z}^2$.

Based on ψ_1 and ψ_2, filters v_j and $w_{j,\ell}^{(d)}$ can be found so that

$$\langle f, \psi_{j,\ell,k}^{(d)}\rangle = f * (v_j * w_{j,\ell}^{(d)})[k]$$

(see [33] for details). To simplify the notation, we suppress the superscript (d) and absorb the distinction between $d = 0$ and 1 by re-indexing the parameter ℓ so that it has double the cardinality. Figure 3 illustrates an example of the filters $(v_j * w_{j,\ell})$.

IV. Constrained TV from a Shearlet Domain

For $\tau \in \mathbb{R}^+$, define the threshold function $T_\tau(x)$ to be x if $|x| \geq \tau$ and zero otherwise. A denoised estimate $\tilde{u}$ from the discrete shearlet transform can expressed as

$$\sum_{M_1}\langle u, \psi_{j,\ell,k}\rangle\psi_{j,\ell,k} + \sum_{M_2}T_\tau(\langle u, \psi_{j,\ell,k}\rangle)\psi_{j,\ell,k}.$$

To obtain a good estimate, the thresholding parameter τ is dependent on j and ℓ. In particular, if we assume the image is subjected to white Gaussian noise, we set $\tau_{j,\ell} = c_j\sigma_{j,\ell}$, where $\sigma_{j,\ell}$ is the standard deviation of the noise at scale j with shearing direction ℓ, and c_j is a scaling parameter. Denote by M_2^C the set of indices of M_2 in the shearlet domain that correspond to the coefficients that would be set to zero in the above reconstruction. A projection operator P_S onto the reconstruction from these coefficients can then be expressed as

$$P_S(u) = \sum_{j,\ell,k \in M_2^C}\langle u, \psi_{j,\ell,k}\rangle\psi_{j,\ell,k}.$$

Our proposed method will be to essentially solve

$$\frac{\partial u}{\partial t} = \nabla \cdot \left(\frac{\phi'(\|\nabla P_S(u)\|)}{\|\nabla P_S(u)\|}\nabla P_S(u)\right) - \lambda_{x,y}(u - u_0)$$

with the Neumann boundary condition $\frac{\partial u}{\partial n} = 0$ on $\partial\Omega$ and the initial condition $u(x, y, 0) = u_0(x, y)$ for $x, y \in \Omega$. The quantity $\lambda_{x,y} = \lambda(x, y)$ is a spatially varying fidelity term based on a measure of local variances that is updated after a number of iterations or progressions of artifical time steps. This adaptive element is intended to improve the recovery of textures by locally controlling the amount of denoising over various image regions according to their content (see [16] for more details).

In an effort to deal with images described as a cartoon-like image plus a residual containing textures and noise, we propose a solution based on a hierarchy of scales. First, we find a solution to (1) which is a cartoon-like denoised image u_c that misses textures and small details. Using the residual image $u_r = u_0 - u_c$, a local variance is calculated as

$$P_{u_r}(i, j) = \frac{1}{mn\sigma^4}\sum_{\tilde{i},\tilde{j}}(u_r(\tilde{i}, \tilde{j}) - \mu[u_r])^2 g_{i,j}(\tilde{i}, \tilde{j}),$$

where $m \times n$ is the size of the image u_0, μ and σ are the mean and the standard deviation of the noise, respectively, and $g_{i,j}(\tilde{i}, \tilde{j}) = g(|i - \tilde{i}|, |j - \tilde{j}|)$ is a normalized and radially symmetric smoothing window, e.g. a Gaussian function $e^{-[(i-\tilde{i})^2 + (j-\tilde{j})^2]/2\beta}$, with $\beta > 0$.

The proposed algorithm can be roughly described as follows, where Δt denotes the time step and $\eta(u) = \nabla \cdot \left(\frac{\phi'(\|\nabla u\|)}{\|\nabla u\|}\nabla u\right)$.

Shearlet TV Algorithm

$$u \qquad\qquad |u - \tilde{u}_C| \qquad\qquad |u - \tilde{u}_S|$$
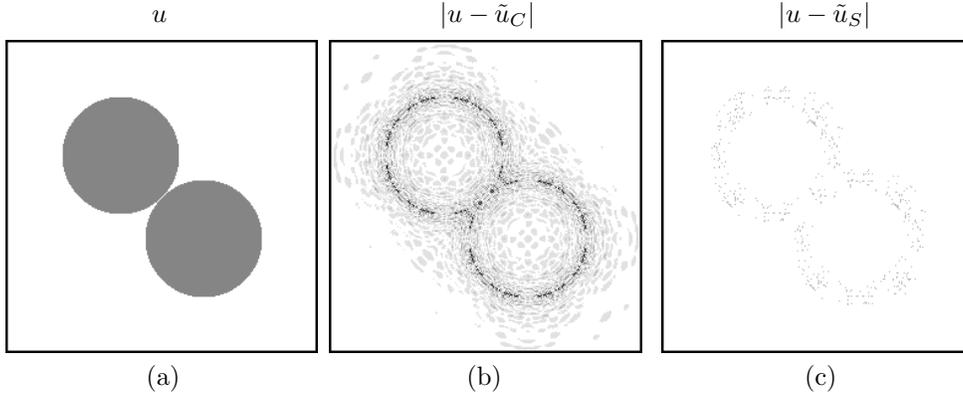

(a)　　　　　　　　　　　　(b)　　　　　　　　　　　　(c)

Fig. 4. Types of residual artifacts after using a thresholding based on the assumption the standard deviation of the noise was 10: (a) Given image. (b) Difference between image (a) and estimate based on the curvelet transform. (c) Difference between image (a) and estimate based on the time-domain shearlet transform.

Given $\epsilon_N > 0$, and a number L of iterations:
$u^0 = u_0, \ s^1 = u_0, \ s^2 = 0, \ \lambda_{x,y} = 0$
while $\frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} |s_{i,j}^2 - s_{i,j}^1| > \epsilon_N$
　　　$s^1 = u^0$
　　　for $k = 0 : L - 1$
　　　　　$u^{k+1} = u^k + \Delta t [\eta(P_S(u^k)) - \lambda_{x,y}(u^k - u_0)]$
　　　end
　　　$\lambda_{x,y} = \eta(u^L)(u^L - u_0)P_{u_r} * g$
　　　$s^2 = u^L$
　　　$u^0 = u^L$
end

In our numerical implementation, we use throughout $L = 7$ and the auxiliary function $\phi(x) = \sqrt{x^2 + \alpha}$ for $\alpha \in \mathbb{R}^+$. We discretize the partial derivatives by means of a centered-difference approximation (e.g. $(\partial u / \partial y)_{i,j} = 1/(2h)(u_{i-1,j} - u_{i+1,j})$ and $(\partial^2 u / \partial y^2)_{i,j} = 1/h^2(u_{i-1,j} - 2u_{i,j} - u_{i+1,j})$ for some $h \in \mathbb{R}^+$).

An alternative method we suggest is based on applying the projection operator to the TV kernel, by solving

$$\frac{\partial u}{\partial t} = P_S \left[\nabla \cdot \left(\frac{\phi'(\|\nabla u\|)}{\|\nabla u\|} \nabla u \right) \right] - \lambda_{x,y}(u - u_0)$$

with the Neumann boundary condition $\frac{\partial u}{\partial n} = 0$ on $\partial\Omega$ and the initial condition $u(x,y,0) = \tilde{u}(x,y)$ for $x,y \in \Omega$. To obtain a slight improvement in performance, we can use $u_r = u^0 - w(u^0)$, where w is an adaptive Wiener filtering operator. As we shall see, this alternative method can be more effective at denoising images in some particular cases.

A related approach is to solve the diffusion equation

$$\frac{\partial u}{\partial t} = \nabla \cdot (\rho(\|\nabla P_C u\|)\nabla P_C u)$$

with periodic boundary conditions and the initial condition $u(x,y,0) = u_0(x,y)$ for $x,y \in \Omega$. The projection operator P_C used in [19] is based on the curvelet transform instead of the shearlet transform. Although both the curvelet and shearlet transforms have the same decay rates, their implementations are significantly different. The discrete shearlet transform can be implemented with no subsampling and with small finitely supported filters in the time-domain. (An example showing a comparison between a frequency based implementation of shearlets and a time-domain implementation is shown in [33].) The residual artifacts for the two transforms after thresholding are shown in Figure 4.

Based on the differences in performance between the transforms, we propose solving

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

with the Neumann boundary condition $\frac{\partial u}{\partial n} = 0$ on $\partial\Omega$ and the initial condition $u(x,y,0) = u_0(x,y)$ for $x,y \in \Omega$.

V. Experimental Results

Numerous experiments on the new methods have been done using the images shown in Figure 5. We have compared our proposed shearlet methods against the curvelet diffusion routine of [19] and several standard estimation routines using total variation, diffusion with $\rho(x) = 1/(1 + x^2/\gamma^2)$ for $\gamma \in \mathbb{R}^+$, and the stationary wavelet transform (SWT) implemented with the Daubechies-Antonini 7/9 filters [18].

Furthermore, we made a comparison with the curvelet diffusion against the shearlet diffusion where the algorithms are essentially the same except for the use of different transforms, the difference being in the boundary assumptions. The results show that the discrete shearlet transform yields improved performance and a significant reduction in the number iterations. Finally, we included a comparison with a standard estimate from shearlets.

Our proposed methods allow one to determine a stopping criterion based on the differences between updates. Since the curvelet diffusion algorithm demonstrated in [19] was iterated until peak performance was obtained, we did the same for all iterative methods presented. All tested algorithms had their free parameters fixed throughout the experiments and were chosen to achieve their best performance over a wide collection of images. All images were processed with equivalent parameters, such as the time step, which was set to be 0.2.

We have found that in our proposed algorithm the use of $u^0 = \tilde{u}$ instead of u_0 yields a slightly better performance. For example, in the experiment with a noisy Barbara image whose SNR is 7.48 dB, in 1 iteration of the adaptive scheme (which corresponds to choosing $L = 7$), when using $u^0 = u_0$ we get an estimate whose SNR is 13.75 dB, whereas, using $u^0 = \tilde{u}$, we get an estimate whose SNR is 14.07 (see Table I). Notice that the implementation of the curvelet-based diffusion scheme in [19] initiates the iteration with $u_0 - \tilde{u}$, which amounts to using $\tilde{u}$ in the first iteration.

Tables I and II display additional results of some of our comparisons. Our other experiments are shown in Figures 6,7,8, and 9.

VI. Conclusion

We have proposed models for implementing TV schemes and a diffusion scheme restricted from a shearlet domain. In particular, we have suggested finding the steady-state solution to an associated Euler-Lagrange equation where only the non-thresholded shearlet coefficients are changed by means of a projection operator. Theoretically, shearlets can represent well images with edges and good denoised estimates can be achieved by adapting in our iteration schemes a sequence of spaces associated with a hierarchy

Fig. 5.   From top left, clockwise: Lily, Barbara, Plane, Elaine.

TABLE I - Routines (performance in PSNR)
Number of iterations used are indicated in parenthesis

Noisy	Shear. TV	Shear. Diff	Curve. Diff
Barbara			
9.98 dB	15.95 (1)	15.70 (6)	14.89 (6)
7.48 dB	14.07 (1)	13.56 (6)	13.00 (9)
5.54 dB	12.10 (4)	11.61 (8)	11.54 (13)
Elaine			
9.75 dB	16.93 (2)	16.73 (16)	16.35 (57)
7.25 dB	16.34 (2)	16.05 (16)	15.72 (80)
5.31 dB	15.63 (2)	15.41 (18)	15.16 (103)

of scales that are spatially localized. The algorithms suggested prove themselves both in terms of efficiency (number of iterations) and performance (quality of the reconstruction) and offer a new paradigm in the development of TV-based algorithms. We intend to address in future research to what extent this new paradigm could be useful for solving problems related to inpainting.

VII. Acknowledgment

References

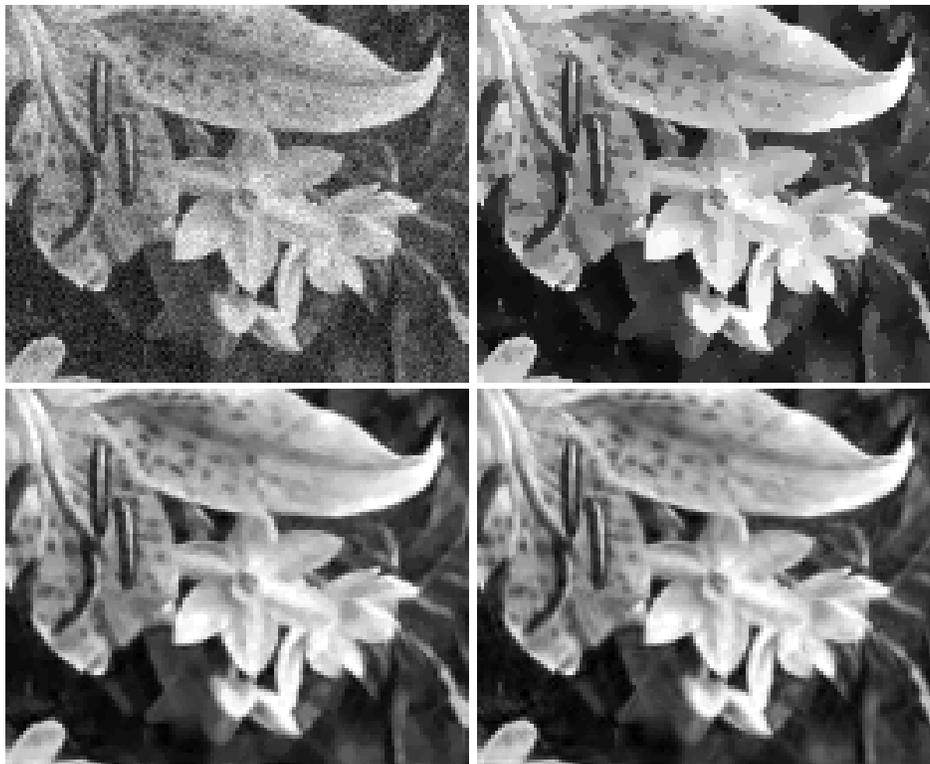[1] S. Mallat, *A Wavelet Tour of Signal Processing*, Academic Press, New York, 1998.

Fig. 7. From top left, clockwise: Noisy image (PSNR=12.69 dB), diffusion based estimate using 37 iterations (SNR=15.96 dB), shearlet based adaptive TV estimate using 2 iterations ($L = 7$, SNR=17.37 dB), shearlet based adaptive TV using a post projection using 23 iterations (SNR=17.45 dB).

TABLE II - Standard Routines (performance in PSNR)
Number of iterations used are indicated in parenthesis

Noisy	TV	Diffusion	SWT
Barbara			
9.98 dB	12.41 (62)	12.87 (34)	13.71
7.48 dB	10.45 (118)	10.96 (60)	11.52
5.54 dB	8.91 (138)	9.57 (96)	9.30
Elaine			
9.75 dB	13.51 (87)	15.75 (51)	15.62
7.25 dB	11.38 (121)	14.65 (88)	12.68
5.31 dB	9.64 (149)	13.70 (137)	9.92

[2] L. Sendur and I. W. Selesnick, "Bivariance shrinkage with local variance estimator", *IEEE Signal Proc. Letters*, **9**(12) (2002), 438–441.

[3] E. J. Candès and D. L. Donoho, "Curvelets A surprisingly effective nonadaptive representation for objects with edges," *in Curves and Surface Fitting, A. Cohen, C. Rabut, and L. L. Schumaker, Eds. Nashville, TN: Vanderbilt Univ. Press*, (2000), 105-120.

[4] J. L. Starck, E. J. Candès, and D. L. Donoho, "The curvelet transform for image denoising", *IEEE Trans. Image Proc.*, **11** (2002), 670-684.

[5] E. J. Candès and D. L. Donoho, "New tight frames of curvelets and optimal representations of objects with

piecewise singularities," *Commun. Pure Appl. Math.*, **57**(2) (2004), 219-266.

[6] E. J. Candès, L. Demanet, D. L. Donoho, and L. Ying, "Fast discrete curvelet transforms," *Multiscale Model. Simul.*, **5**(3) (2006), 861-899.

[7] G. R. Easley, D. Labate, and W.-Q. Lim, "Optimally sparse image representations using shearlets", *Proc. 40th Asilomar Conf. on Signals, Systems, and Computers*, Monterey (2006).

[8] D. Labate, and W.-Q. Lim, G. Kutyniok, and G. Weiss, "Sparse multidimensional representation using shearlets", Wavelets XI (San Diego, CA, 2005), 254–262, *SPIE Proc.*, **5914**, SPIE, Bellingham, WA, 2005.

[9] 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.

[10] R. R. Coifman and A. Sowa, "Combining the calculus of variations and wavelets for image enhancement," *Appl. Comput. Harmon. Anal.*, **9** (2000), 1-18.

[11] S. Durand and J. Froment, "Reconstruction of wavelet coefficients using total variation minimization", *SIAM J. Sci. Comput.*, **24**(5) (2003), 1754–1767.

[12] J. Ma and M. Fenn, "Combined complex ridgelet shrinkage and total variation minimization", *SIAM*

Fig. 8. From top left, clockwise: Noisy image (SNR=7.25 dB), shearlet based diffusion estimate using 16 iterations (SNR=16.05 dB), shearlet based adaptive TV estimate using 2 iterations ($L = 7$, SNR=16.34 dB), curvelet based diffusion estimate using 80 iterations (SNR=15.72).

J. Sci. Comput., **28**(3) (2006), 984-1000.

[13] 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.

[14] G. R. Easley, and F. Colonna. "Improvements on total variation algorithms for denoising", preprint, (2008).

[15] D. Goldfarb, S. Osher, and W. Yin "Image cartoon-texture decomposition and feature selection using the total variation regularized L1 functional", In *Variational, Geometric, and Level Set Methods in Computer Vision, Lecture Notes in Computer Science*, Springer **3752**, (2005) 73–84.

[16] G. Gilboa, Y. Y. Zeevi, and N. Sochen, "Texture preserving variational denoising using an adaptive fidelity term", *Proc. VLSM*, Nice (2003), 137–144.

[17] E. Tadmor, S. Nezzar, and L. Vese, "A multiscale image representation using hierarchical (BV, L^2) decomposition", *Technical Report CAM03-20*.

[18] M. Antonini, M. Barlaud, P. Mathieu and I. Daubechies. "Image coding using wavelet transform. " *IEEE Trans. Image Proc.*, **1**, (1992), 205–220.

[19] J. Ma and G. Plonka, "Combined curvelet shrinkage and nonlinear anisotropic diffusion", *IEEE Trans. Image Proc.*, **16** (2007), 2198–2206.

[20] L. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithm", *Physica D*, **60**(1992), 259–268.

[21] T. Chan, G. Golub, and P. Mulet, "A non-linear primal-dual method for total variation-based image restoration", *SIAM J. Sci. Comp.*, **20**(1999), 1964–1977.

[22] J. Carter, "Dual methods for total variation-based image restoration", Doctoral Dissertation, University of California, Los Angeles, CA, 2001.

[23] A. Chambolle, "An algorithm for total variation minimization and applications", *J. Math. Imaging Vision*, **20** (2004), 89–97.

[24] 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.

[25] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion", *IEEE Trans. Pattern Anal. Mach. Intel.*, **12** (1990), 629–639.

[26] S. Yi, D. Labate, G. R. Easley, and H. Krim, "A shearlet approach to edge detection and analysis", to appear in *Proc. IEEE ICIP* (2008).

Fig. 6. From top left, clockwise: Noisy image (SNR=7.48 dB), shearlet based adaptive TV estimate using a post projection with 2 iterations (SNR=13.90 dB), shearlet based adaptive TV estimate using 1 iteration ($L = 7$, SNR=14.07 dB), curvelet based diffusion using 9 iterations (SNR=13.00).
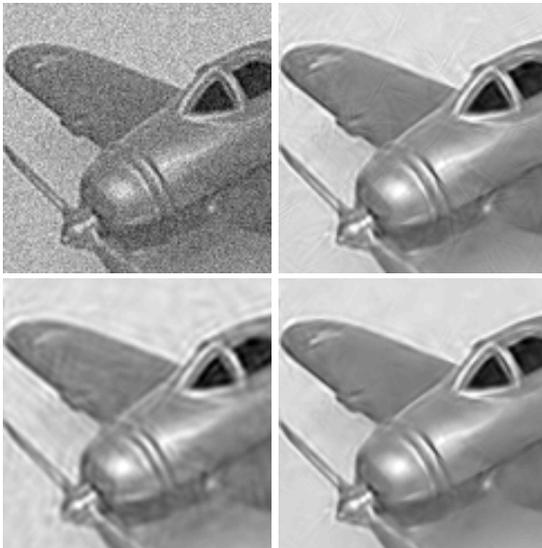

Fig. 9. From top left, clockwise: Noisy image (SNR=6.63 dB), shearlet based estimate (SNR=17.51 dB), shearlet based adaptive TV estimate using 1 iteration ($L = 7$, SNR=18.36 dB), curvelet based diffusion estimate using 56 iterations (SNR=14.80).

[27] Y. Meyer, *Wavelets and Operators*, Cambridge Stud. Adv. Math. **37**, Cambridge Univ. Press, Cambridge, UK, 1992.

[28] G. Kutyniok and D. Labate, "Resolution of the wavefront set using continuous shearlets", to appear in *Trans. Amer. Math. Soc.* (2008).

[29] K. Guo, D. Labate and W. Lim, "Edge analysis and identification using the continuous shearlet transform", preprint (2008).

[30] K. Guo and D. Labate, "Optimally Sparse multidimensional representation using shearlets", *SIAM J. Math. Anal.* **39** (2007), 298–318.

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

[32] A. L. Cunha, J. Zhou, M. N. Do, "The nonsubsampled contourlet transform: Theory, design, and applications", *IEEE Trans. Image Proc.*, **15** (2006), 3089–3101.

[33] G. R. Easley, D. Labate, and W. Lim. "Sparse directional image representations using the discrete shearlet transform", to appear in *Appl. Comput. Harmon. Anal.* (2008).

[34] P. Mrzek, J. Weickert, and G. Steidl, "Diffusion inspired shrinkage functions and stability results for wavelet denoising", *Int. J. Comput. Vis.*, **64** (2005), 171-186.