Math. Model. Nat. Phenom.
Vol. 8, No. 1, 2013, pp. 60–74
DOI: 10.1051/mmnp/20138104

3D Data Denoising Using Combined Sparse Dictionaries

G. Easley[1], D. Labate[2] *, P. Negi[2]
[1] System Planning Corporation, Arlington, VA 22201, USA
[2] Department of Mathematics, University of Houston, Houston, Texas 77204, USA

Abstract. Directional multiscale representations such as shearlets and curvelets have gained increasing recognition in recent years as superior methods for the sparse representation of data. Thanks to their ability to sparsely encode images and other multidimensional data, transform-domain denoising algorithms based on these representations are among the best performing methods currently available. As already observed in the literature, the performance of many sparsity-based data processing methods can be further improved by using appropriate combinations of dictionaries. In this paper, we consider the problem of 3D data denoising and introduce a denoising algorithm which uses combined sparse dictionaries. Our numerical demonstrations show that the realization of the algorithm which combines 3D shearlets and local Fourier bases provides highly competitive results as compared to other 3D sparsity-based denosing algorithms based on both single and combined dictionaries.

Keywords and phrases: curvelets, denoising, nonlinear approximations, pursuit algorithms, shearlets, sparse approximations, wavelets

Mathematics Subject Classification: 42C15, 42C40

1. Introduction

The notion of sparse representations has been one of the most pervasive and seminal concepts in the applied harmonic analysis and signal processing literature during the last decade. Intuitively, a representation is *sparse* if it allows one to efficiently approximate functions in a certain class using only a "small" number of terms when functions are expanded with respect to the representation.

Sparsity can be exploited very effectively in denoising applications. In fact, constructing sparse representations for functions in a certain class entails the ability to efficiently capture their fundamental structures, hence enabling to extract the features of interest and remove the unwanted components. A transform-domain denoising algorithm exploits the concept of sparsity through the assumption that the true data are well approximated by a linear combination of few basis elements with respect to an appropriate representation. Hence, by preserving the few large-magnitude representation coefficients and discarding the rest, the true signal can be effectively estimated. This concept was formalized through the celebrated shrinkage algorithms originally introduced by Donoho and Johnstone [7, 10–12], which have been spectacularly successful in the signal and image processing community during the past 15 years.

*Corresponding author. E-mail: dlabate@math.uh.edu

© EDP Sciences, 2013

Article published by EDP Sciences and available at http://www.mmnp-journal.org or http://dx.doi.org/10.1051/mmnp/20138104

To illustrate this denoising strategy in the classical setting of additive white Gaussian noise, let us suppose that we want to recover $f \in L^2(\mathbb{R}^d)$ from the noisy observations

$$y = f + n, \tag{1.1}$$

where n is zero-mean white Gaussian noise with variance σ^2. If the function f to be recovered is known to be sparse under an appropriate dictionary, say, orthogonal wavelets, then the classical shrinkage algorithm offers a simple yet very effective approach. This consists in: (i) applying the transform analysis operator, denoted by $\mathcal{W}$, to the noisy data, (ii) applying a thresholding operator $\mathcal{T}$ and, finally, (iii) applying the inverse transform operator to obtain an estimator $\tilde{f}$ of f. That is,

$$\tilde{f} = \mathcal{W}^{-1}\,\mathcal{T}\,\mathcal{W}\,f. \tag{1.2}$$

The performance of this estimation is closely related to the sparsity of the representation system. In fact, suppose that $\mathcal{W}$ is unitary and $\mathcal{T}$ is the soft thresholding operator, defined by $\mathcal{T}_\tau(\alpha) = \text{sign}\,(\alpha)(|\alpha| - \tau)_+$, where τ is a fixed parameter. In this case, Donoho and Johnstone [10] have shown that, if the m-term approximation error decays as

$$\|f - f_m\|_{L^2}^2 \leq C\, m^{-k}, \qquad \text{as } m \to \infty,$$

where f_m is the m-term approximation of f, then there is a constant $C' > 0$ so that

$$E\left(\|f - \tilde{f}\|_{L^2}^2\right) \leq C'\, \sigma^{\frac{2k}{k+1}}, \qquad \text{as } \sigma \to 0,$$

where E is the expectation and σ is the standard deviation of the noise. For example, curvelets and shearlets are optimally efficient for representing smooth images with edges, so that curvelet-based and shearlet-based shrinkage algorithms are extremely effective for this type of data.

However, data found in applications are usually complex and there is no single representation system that can optimally approximate all the features of interest. While curvelets and shearlets are very efficient for representing data with distributed singularities such as edges, Fourier bases are better at dealing with oscillatory features, usually described as *texture*. Therefore, it seems reasonable to consider combined representations made up of many different types of waveforms, each type having its own area of specialization. The idea of coding images as a multi-layered representation of a piecewise smooth layer and a texture layer is rather old [22, 28], but only more recently the concept of combined representations has been considered within the setting of sparse data decompositions. In particular, one emerging approach is based on the concept *morphological diversity* (see discussion in [31, 33]), which assumes that data can be modeled as a sum:

$$x = \sum_{k=1}^{K} x_k,$$

where each *morphological components* x_k of the data is sparse with respect to a given dictionary Φ_k. In addition, the various dictionaries satisfy a form of *incoherence*, that is, each components x_k of the data is sparse with respect to a dictionary Φ_k, but not with respect to the other dictionaries Φ_ℓ, for $\ell \neq k$.

In this paper, we have focussed on 3D data denoising, which is an area of increasing interest for video restoration and preprocessing of biological data. For this task, some of the most competitive algorithms currently available are transform-domain algorithms which rely on the sparsity of advanced directional multiscale systems such as curvelets, shearlets and surfacelets [25, 29, 34]. As suggested above and as indicated by several recent results in the literature, the performance of many sparsity-based algorithm can be further improved by considering dictionaries which are obtained as appropriate combinations of several representations, each one specialized in different features of the data to be analyzed (e.g., textures and piecewise smooth regions). Following the general philosophy of morphological diversity, we have assumed that typical 3D data of interest (e.g. videos of natural scenery, biological data) are

https://doi.org/10.1051/mmnp/20138104 Published online by Cambridge University Press

well represented through combined dictionaries based on wavelets, curvelets, shearlets and Fourier bases. Using this approach we have derived highly performing denoising algorithms which were validated through extensive numerical experiments. As a benchmark, our results were compared against standard and state-of-the-art denoising algorithms based on both single and combined dictionaries, showing that, in particular, the realization of our algorithm which combines 3D shearlets and local Fourier bases provides the best performance in terms of noise removal.

The paper is organized as follows. In Section 2 we will briefly survey the theory of sparse representations and recall the construction of 3D shearlets, which are especially efficient for the representation of piecewise smooth functions of three variables. In Section 3 we present an algorithm for 3D data denoising which uses an appropriate combination of sparse representations. Finally, in Section 4 we present several numerical demonstrations of our denoising algorithm, and compare it against other standard and state-of-the-art denoising algorithms.

2. Sparse multidimensional representations

Sparsity and sparse representations are leading concepts not only in the current signal and image processing literature, but also in a variety of areas such as inverse problems, numerical analysis and data mining. We will not attempt to comprehensively review this topic and its applications, but will only focus on the aspects which are most closely related to the problem of data denoising. We start with some basic definitions.

2.1. Sparse representations: beyond wavelets

Let $\mathcal{B} = \{b_i : i \in I\}$ be an orthonormal basis or even a Parseval frame for a subset U of $L^2(\mathbb{R}^d)$, and, for $f \in U$, let us consider its expansion in $\mathcal{B}$, that is, $f = \sum_{i\in I} \alpha_i(f)\, b_i$. The system $\mathcal{B}$ is said to provide a *sparse representation* of U if the sorted magnitudes of the expansion coefficients $|\alpha(f)|_{|(i)|}$ have rapid decay according to the power law

$$|\alpha(f)|_{|(N)|} \leq C\, \|f\|_{L^2}\, N^{-s}, \quad \text{as } N \to \infty,$$

where $C > 0$ is a constant independent of f. This implies that, letting f_N be the approximation of f obtained by using the N largest coefficients of the expansion of f in $\mathcal{B}$, the *nonlinear approximation error*, defined by $\|f - f_N\|^2_{L^2}$, decays as

$$\|f - f_N\|^2_{L^2} \leq C\, \|f\|^2_{L^2}\, N^{1-2s}, \quad \text{as } N \to \infty.$$

For example, wavelet bases provide sparse representations for piecewise smooth functions of one variable. In fact, if f is uniformly Lipschitz γ, with $\gamma > 1/2$, except for finitely many discontinuities on $[0, 1]$, then one can show[1] that the sorted magnitudes of the wavelet coefficients decay as

$$|\alpha(f)|_{|(N)|} \leq C\, \|f\|_{C^\gamma}\, N^{-\gamma+\frac{1}{2}}, \quad \text{as } N \to \infty.$$

Hence, denoting by $f_N^{(w)}$ the nonlinear approximation of f using a wavelet basis, the nonlinear approximation error decays as

$$\|f - f_N^{(w)}\|_2^2 \leq C\, \|f\|^2_{C^\gamma}\, N^{-2\gamma}, \quad \text{as } N \to \infty. \tag{2.1}$$

This is the optimal rate achievable for this class of functions, whereas, by contrast, approximations using Fourier bases can only decay as fast as $O(N^{-2})$, due to the presence of the discontinuities.

However, while wavelets are particularly effective when approximating functions with point discontinuities, it is not difficult to show that they do not provide very sparse representations for multi-variable functions containing discontinuities along curves or other multidimensional manifolds. Consider, for

[1] One need to assume that the wavelet generator has sufficiently many vanishing moments (cf. [27, Ch.9]).

https://doi.org/10.1051/mmnp/20138104 Published online by Cambridge University Press

example, functions of two variables of the form $f = \chi_\Omega$, where Ω is a bounded subset of $\mathbb{R}^2$ whose boundary has finite length. Then a simple calculation (cf. [27, Sec. 9.3.1]) shows that, in this case, the sorted magnitudes of the wavelet coefficients only decay as

$$|\alpha(f)|_{|(N)|} \leq C\, N^{-1}, \quad \text{as } N \to \infty, \tag{2.2}$$

which is below the optimal theoretical rate of $O(N^{-3/2})$ [8].

The reason for the limitations of wavelets is their intrinsic isotropic nature, which makes these systems not sufficiently flexible to capture the dominant features of multivariable functions efficiently. Starting with the introduction of wedgelets [9] and ridgelets [3], it was recognized that to achieve sparser representations of multidimensional data, one has to consider larger representation systems offering the ability to handle efficiently the anisotropic features which dominate most classes of multivariate functions. The construction of curvelets [4] has been a milestone in this area of investigation, by showing that it is possible to define non-adaptive representations providing (nearly) optimal sparse representations for a large class of functions of two variables. Curvelets form a tight frame of functions defined not only at various scales and locations, like wavelets, but also at various orientations, with the number of orientations increasing at finer scales and with highly anisotropic supports. Following the general idea of the curvelets, several other directional multiscale systems were introduced, most notably the *shearlets*, originally introduced by one of the authors and their collaborators in [17,24]. Unlike curvelets, shearlets exploit the general framework of affine systems and use shearing matrices rather than rotations to control the directional selectivity of the system. As a result, shearlets provide a very flexible approach for the construction of sparse representations in any dimensions and a unified framework for both the continuous and discrete setting. Shearlets are currently the only systems known to provide nearly optimally sparse representations for a large class of functions of two and three variables [18–20].

2.2. 3D shearlet representations

A 3D shearlet system is obtained by combining three function systems associated with the pyramidal regions

$$\mathcal{P}_1 = \left\{(\xi_1,\xi_2,\xi_3) \in \mathbb{R}^3 : |\tfrac{\xi_2}{\xi_1}| \leq 1, |\tfrac{\xi_3}{\xi_1}| \leq 1\right\}, \mathcal{P}_2 = \left\{(\xi_1,\xi_2,\xi_3) \in \mathbb{R}^3 : |\tfrac{\xi_1}{\xi_2}| < 1, |\tfrac{\xi_3}{\xi_2}| \leq 1\right\},$$

and $\mathcal{P}_3 = \left\{(\xi_1,\xi_2,\xi_3) \in \mathbb{R}^3 : |\tfrac{\xi_1}{\xi_3}| < 1, |\tfrac{\xi_2}{\xi_3}| < 1\right\},$

in which the Fourier space $\widehat{\mathbb{R}}^3$ is partitioned (see Fig. 1). To define such systems, let ϕ be a C^∞ univariate function such that $0 \leq \hat{\phi} \leq 1$, $\hat{\phi} = 1$ on $[-\frac{1}{16}, \frac{1}{16}]$ and $\hat{\phi} = 0$ outside the interval $[-\frac{1}{8}, \frac{1}{8}]$. That is, ϕ is the scaling function of a Meyer wavelet. For $\xi = (\xi_1,\xi_2,\xi_3) \in \widehat{\mathbb{R}}^3$, define

$$\widehat{\Phi}(\xi) = \widehat{\Phi}(\xi_1,\xi_2,\xi_3) = \hat{\phi}(\xi_1)\,\hat{\phi}(\xi_2)\,\hat{\phi}(\xi_3) \tag{2.3}$$

and let $W(\xi) = \sqrt{\widehat{\Phi}^2(2^{-2}\xi) - \widehat{\Phi}^2(\xi)}$. It follows that

$$\widehat{\Phi}^2(\xi) + \sum_{j\geq 0} W^2(2^{-2j}\xi) = 1 \ \text{ for } \xi \in \mathbb{R}^3. \tag{2.4}$$

Note that each function $W_j = W(2^{-2j}\,\cdot)$, $j \geq 0$, is supported inside the Cartesian corona

$$[-2^{2j-1}, 2^{2j-1}]^3 \setminus [-2^{2j-4}, 2^{2j-4}]^3 \subset \widehat{\mathbb{R}}^3,$$

and the functions W_j^2, $j \geq 0$, produce a smooth tiling of $\widehat{\mathbb{R}}^3$. Next, let $V \in C^\infty(\mathbb{R})$ be such that $V(0) = 1$, $V^{(n)}(0) = 0$, for any $n \geq 1$, $\operatorname{supp} V \subset [-1,1]$ and

$$|V(u-1)|^2 + |V(u)|^2 + |V(u+1)|^2 = 1 \quad \text{for } |u| \leq 1. \tag{2.5}$$

https://doi.org/10.1051/mmnp/20138104 Published online by Cambridge University Press

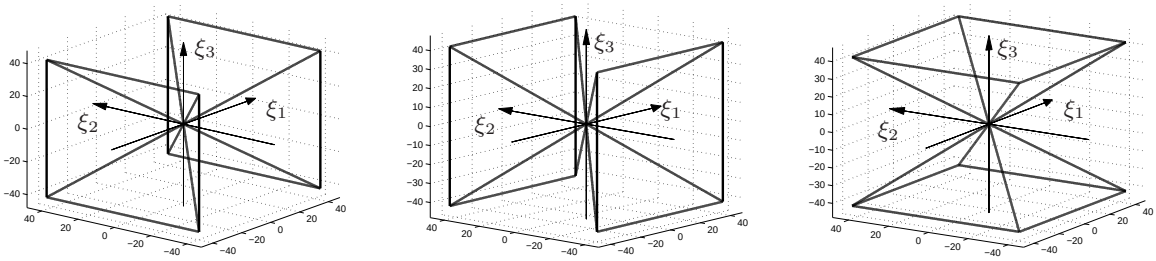


FIGURE 1. From left to right, the figure illustrates the pyramidal regions $\mathcal{P}_1$, $\mathcal{P}_2$ and $\mathcal{P}_3$ in the frequency space $\widehat{\mathbb{R}}^3$.

It was shown in [18] that there are several examples of functions V satisfying these properties.

For $d = 1, 2, 3$, $\ell = (\ell_1, \ell_2) \in \mathbb{Z}^2$, the 3D *shearlet systems associated with the pyramidal regions* $\mathcal{P}_d$ are defined as the collections

$$\{\psi^{(d)}_{j,\ell,k} : j \geq 0, -2^j \leq \ell_1, \ell_2 \leq 2^j, k \in \mathbb{Z}^3\}, \tag{2.6}$$

where

$$\hat{\psi}^{(d)}_{j,\ell,k}(\xi) = |\det A_{(d)}|^{-j/2}\, W(2^{-2j}\xi)\, F_{(d)}(\xi A^{-j}_{(d)} B^{[-\ell]}_{(d)})\, e^{2\pi i \xi A^{-j}_{(d)} B^{[-\ell]}_{(d)} k}, \tag{2.7}$$

$F_{(1)}(\xi_1, \xi_2, \xi_3) = V(\frac{\xi_2}{\xi_1})V(\frac{\xi_3}{\xi_1})$, $F_{(2)}(\xi_1, \xi_2, \xi_3) = V(\frac{\xi_1}{\xi_2})V(\frac{\xi_3}{\xi_2})$, $F_{(3)}(\xi_1, \xi_2, \xi_3) = V(\frac{\xi_1}{\xi_3})V(\frac{\xi_2}{\xi_3})$, the anisotropic dilation matrices $A_{(d)}$ are given by

$$A_{(1)} = \begin{pmatrix} 4 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix},\ A_{(2)} = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 2 \end{pmatrix},\ A_{(3)} = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 4 \end{pmatrix},$$

and the *shear matrices* are defined by

$$B^{[\ell]}_{(1)} = \begin{pmatrix} 1 & \ell_1 & \ell_2 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix},\ B^{[\ell]}_{(2)} = \begin{pmatrix} 1 & 0 & 0 \\ \ell_1 & 1 & \ell_2 \\ 0 & 0 & 1 \end{pmatrix},\ B^{[\ell]}_{(3)} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \ell_1 & \ell_2 & 1 \end{pmatrix}.$$

Due to the assumptions on W and v, the elements of the system of shearlets (2.6) are well localized and bandlimited. In particular, the shearlets $\hat{\psi}^{(1)}_{j,\ell,k}(\xi)$ can be written more explicitly as

$$\hat{\psi}^{(1)}_{j,\ell_1,\ell_2,k}(\xi) = 2^{-2j}\, W(2^{-2j}\xi)\, V\Big(2^j \frac{\xi_2}{\xi_1} - \ell_1\Big)\, V\Big(2^j \frac{\xi_3}{\xi_1} - \ell_2\Big)\, e^{2\pi i \xi A^{-j}_{(1)} B^{[-\ell_1, -\ell_2]}_{(1)} k}, \tag{2.8}$$

showing that their supports are contained inside the trapezoidal regions

$$\{(\xi_1, \xi_2, \xi_3) : \xi_1 \in [-2^{2j-1}, -2^{2j-4}] \cup [2^{2j-4}, 2^{2j-1}], |\frac{\xi_2}{\xi_1} - \ell_1 2^{-j}| \leq 2^{-j}, |\frac{\xi_3}{\xi_1} - \ell_2 2^{-j}| \leq 2^{-j}\}.$$

This expression shows that these support regions become increasingly more elongated at fine scales, due to the action of the anisotropic dilation matrices $A^j_{(1)}$, with the orientations of these regions controlled by the shearing parameters ℓ_1, ℓ_2. A typical support region is illustrated in Fig. 2. Similar properties hold for the elements associated with the regions $\mathcal{P}_2$ and $\mathcal{P}_3$.

A Parseval frame of shearlets for $L^2(\mathbb{R}^3)$ is obtained by combining the systems of shearlets associated with the 3 pyramidal regions $\mathcal{P}_d$, $d = 1, 2, 3$, together with a coarse scale system, which will take care of

https://doi.org/10.1051/mmnp/20138104 Published online by Cambridge University Press

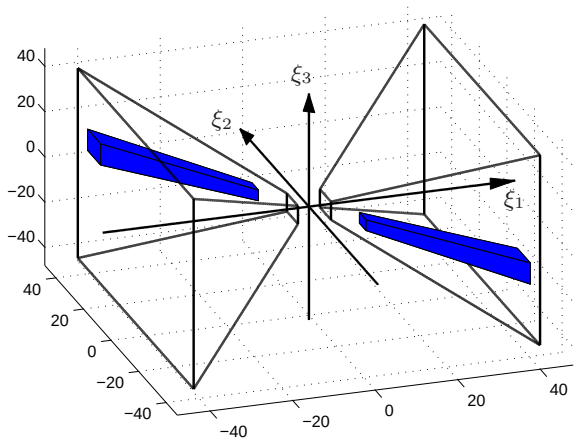


FIGURE 2. *Frequency support of a representative shearlet function $\psi_{j,\ell,k}$, inside the pyramidal region $\mathcal{P}_1$. The orientation of the support region is controlled by $\ell = (\ell_1, \ell_2)$; its shape is becoming more elongated as j increases ($j = 4$ in this plot).*

the low frequency region. Namely, we define the *3D shearlet systems for* $L^2(\mathbb{R}^3)$ as the collections

$$\Big\{\widetilde{\psi}_{-1,k} : k \in \mathbb{Z}^3\Big\} \bigcup \Big\{\widetilde{\psi}_{j,\ell,k,d} : j \geq 0, |\ell_1| < 2^j, |\ell_2| \leq 2^j, k \in \mathbb{Z}^3, d = 1,2,3\Big\}$$
$$\bigcup \Big\{\widetilde{\psi}_{j,\ell,k} : j \geq 0, \ell_1, \ell_2 = \pm 2^j, k \in \mathbb{Z}^3\Big\} \tag{2.9}$$

consisting of:

- the *coarse-scale shearlets* $\{\widetilde{\psi}_{-1,k} = \Phi(\cdot - k) : k \in \mathbb{Z}^3\}$, where Φ is given by (2.3);
- the *interior shearlets* $\{\widetilde{\psi}_{j,\ell,k,d} = \psi^{(d)}_{j,\ell,k} : j \geq 0, |\ell_1||\ell_2| < 2^j, k \in \mathbb{Z}^3, d = 1,2,3\}$, where $\psi^{(d)}_{j,\ell,k}$ are given by (2.7);
- the *boundary shearlets* $\{\widetilde{\psi}_{j,\ell,k,d} : j \geq 0, |\ell_1| \leq 2^j, \ell_2 = \pm 2^j, k \in \mathbb{Z}^3, d = 1,2,3\}$, obtained by appropriately joining together $\psi^{(1)}_{j,\ell,k}$, $\psi^{(2)}_{j,\ell,k}$ and $\psi^{(3)}_{j,\ell,k}$, for $\ell_1, \ell_2 = \pm 2^j$ (see [20,21] for details). Note that these functions are band-limited and smooth in the Fourier domain.

For simplicity of notation we will write the 3D shearlet system (2.9) as

$$\{\widetilde{\psi}_\mu,\ \mu \in \mathcal{M}\}, \tag{2.10}$$

where $\mathcal{M} = \mathcal{M}_C \cup \mathcal{M}_I \cup \mathcal{M}_B$ are the indices associated with the *coarse-scale shearlets,* the *interior shearlets* and the *boundary shearlets*, as given by (2.9). Hence, we have the following result which uses properties (2.4)-(2.5) and whose proof is found in [20].

Theorem 2.1. *The 3D system of shearlets* (2.10) *is a Parseval frame of* $L^2(\mathbb{R}^3)$. *That is:*

$$\sum_{\mu \in \mathcal{M}} |\langle f, \widetilde{\psi}_\mu \rangle|^2 = \|f\|^2, \quad \forall f \in L^2(\mathbb{R}^3).$$

2.3. Optimally sparse 3D approximations using shearlets

One of the most important results about shearlet representations is that the 3D Parseval frames of shearlets $\{\widetilde{\psi}_\mu,\ \mu \in \mathcal{M}\}$ achieve an essentially optimal approximation rate for piecewise smooth functions

https://doi.org/10.1051/mmnp/20138104 Published online by Cambridge University Press

of three variables [19,20]. More precisely, let $\mathcal{D}$ be the class of indicator functions of sets $B \subset [0,1]^3$ whose boundary is C^2 regular and let $C_c^2([0,1]^3)$ be the collection of twice differentiable functions supported inside $[0,1]^3$. Hence, we define the set $\mathcal{E}^2([0,1]^3)$ of *functions which are C^2 away from a C^2 surface* as the collection of functions of the form

$$f = f_0 + f_1\,\chi_B,$$

where $f_0, f_1 \in C_c^2([0,1]^3)$, $B \in \mathcal{D}$ and $\|f\|_{C^2} = \sum_{|\alpha|\leq 2}\|D^\alpha f\|_\infty \leq 1$. This is a reasonable model for piecewise smooth 3D data, extending the similar 2D model of *cartoon-like images* introduced by Donoho in [9]. We have the following result [20].

Theorem 2.2. *Let $f \in \mathcal{E}^2$ and f_N^S be the N–term approximation of f obtained from the N largest coefficients of its shearlet expansion, namely*

$$f_N^S = \sum_{\mu\in I_N} \langle f, \widetilde{\psi}_\mu\rangle\,\widetilde{\psi}_\mu,$$

where $I_N \subset M$ is the set of indices corresponding to the N largest entries of the sequence $\{|s_\mu(f)| = |\langle f, \widetilde{\psi}_\mu\rangle|^2 : \mu \in M\}$. Then

$$\|f - f_N^S\|_2^2 \leq C\,N^{-1}\,(\log N)^2.$$

Up to the log-like factor, Theorem 2.2 yields the optimal approximation rate, outperforming wavelet approximations which only achieve an approximation rate of order $O(N^{-1/2})$. Shearlets are the only system known to satisfy such approximation properties. As a consequence of this result, it follows from the observation of Section 1 that, for the solution of the denoising problem (1.1), under the assumption that $f \in \mathcal{E}^2$, a shearlets-based shrinkage estimator achieves the essentially optimal rate $O(\sigma)$, in contrast to wavelets which only achieve the rate $O(\sigma^{\frac{2}{3}})$. An implementation of the shearlet-based estimator for 3D data denoising has been derived by two of the authors in [29], where several numerical demonstrations have shown that the performance of this algorithm is very competitive, as compared to other advanced multiscale methods including 3D curvelets and surfacelets.

3. 3D data denoising using combined dictionaries

Let us examine the denoising problem from the point of view of discrete signals. Let $x \in \mathbb{R}^{N^3}$ and suppose that we want to recover x from the noisy observations (1.1). Since the signals of interest have some "structure", adopting the point of view of sparse representations we assume that x is sparse in an overcomplete dictionary, represented in a matrix form as $\mathcal{D}$ where $\mathcal{D} \in \mathbb{R}^{N^3\times K}$, with $K \gg N^3$. That is, writing

$$x = \mathcal{D}\alpha = \sum_{k=1}^{K} \alpha_k\,d_k,$$

where $\mathcal{D} = [d_i, \ldots, d_k]$, $d_k \in \mathbb{R}^{N^3}$, we expect that "many" of the representations coefficients α_k are negligible. If we want to minimize the number of non-negligible coefficients, we can set up the minimization problem

$$\hat{\alpha} = \min\|\alpha\|_0 \quad \text{subject to } \|y - \mathcal{D}\alpha\|_2 \leq \sigma, \tag{3.1}$$

where the ℓ_0 norm[2] is counting the number of nonzero entries of $\alpha = (\alpha_k)$, that is, $\|(\alpha_k)\|_0 = \#\{k : \alpha_k \neq 0\}$. Since the algorithmic solution of (3.1) is NP-hard, this problem is usually modified by relaxing ℓ_0 to an ℓ_1-norm, hence leading to the *basis pursuit denoising* (BPDN) problem [5]

$$\hat{\alpha} = \min\|\alpha\|_1 \quad \text{subject to } \|y - \mathcal{D}\alpha\|_2 \leq \sigma, \tag{3.2}$$

[2]This is not technically a norm, but this abuse of notation is customary in the literature.

https://doi.org/10.1051/mmnp/20138104 Published online by Cambridge University Press

and the estimator $\hat{x} = \mathcal{D}\hat{\alpha}$. There are many ways to solve this problem rather efficiently including interior point methods and gradient projections (cf. [15]). For an appropriate Lagrange multiplier λ, the solution of (3.2) is exactly the solution of the unconstrained optimization problem[3]

$$\min_{\alpha} \lambda\|\alpha\|_1 + \frac{1}{2}\|y - \mathcal{D}\alpha\|_2^2. \tag{3.3}$$

In the special case where the analysis operator $\mathcal{W}$ associated with the dictionary $\mathcal{D}$ is unitary, then equation (1.2), with $\mathcal{D} = \mathcal{W}^T$, is the unique solution of problem (3.2) or (3.3). Interestingly, in this case, the shrinkage algorithm also yields the solution of problem (3.1). We refer to Ch. 5 in [15] for further detail. When the dictionary is redundant, then the solution of minimization problem (3.2) or (3.3) is not necessarily the solution of the shrinkage method (1.2). Yet, also in this case, appropriate iterative-shrinkage algorithms have been introduced which extend the classical Donoho-Johnstone wavelet shrinkage method, such as the algorithms introduced by Starck et al. [1,32] and the celebrated algorithm of Daubechies et al. [6].

We are interested in the situation where the data to be recovered are known to be a superposition of several components, each one having a sparse representation with respect to a certain dictionary. In this case, we model the data x to be recovered in (1.1) as $x = \sum_{k=1}^{K} x_k$ and we use a dictionary built by amalgamating several subdictionaries $\mathcal{D}_1, \ldots, \mathcal{D}_K$ such that each x_k has a sparse representation in the subdictionary $\mathcal{D}_k$ but its representation in the subdictionaries $\mathcal{D}_l$, $l \neq k$, is not sparse. Specifically, in many situations it is reasonable to assume that the data contain textured components along with piecewise smooth components. Hence, we assume that x is a superposition

$$x = x_p + x_t,$$

where x_p and x_t are the piecewise smooth component and textured component of the data, respectively. In this setting, for the subdictionary associated with texture component of the data we can choose a local discrete cosine dictionary or an appropriate Gabor frame, which are known to be sparse for locally periodic patterns. For the piecewise smooth component of the data, we can choose a shearlet or a curvelet dictionary, which are known to be sparse for this type of data. The incoherence of the two dictionary has been verified heuristically in [31] (using DCT and curvelet dictionaries) and more recently and rigorously in [23]. Then, to solve this problem, we can set up the following minimization problem:

$$\hat{\alpha}_t, \hat{\alpha}_p = \min_{\alpha_t, \alpha_p} \lambda\left(\|\alpha_t\|_1 + \|\alpha_p\|_1\right) + \frac{1}{2}\|y - \mathcal{D}_t\,\alpha_t - \mathcal{D}_p\,\alpha_p\|_2^2, \tag{3.4}$$

where $\mathcal{D}_t$, $\mathcal{D}_p$ are the dictionary associated with the piecewise smooth component and textured component of the data, respectively.. The final estimate is then found by adding together the two components obtained by $\hat{x}_p = \mathcal{D}_p\hat{\alpha}_p$ and $\hat{x}_t = \mathcal{D}_t\hat{\alpha}_t$. Note that, since the dictionaries are assumed to be tight frames, then $\mathcal{D}_p$ is the Moore-Penrose pseudo inverse of the analysis operator $\mathcal{W}_p$ associated with piecewise smooth data, i.e. $\mathcal{D}_p = \mathcal{W}_p^{\dagger}$ and, similarly, $\mathcal{D}_t$ is the Moore-Penrose pseudo inverse of the analysis operator $\mathcal{W}_t$ associated with texture data, i.e., $\mathcal{D}_t = \mathcal{W}_t^{\dagger}$.

For reason which will become clear below, rather than using a sparsity-based *synthesis model* as in (3.4), we have preferred to use a sparsity-based *analysis model* leading to the minimization problem

$$\hat{x}_p, \hat{x}_t = \arg\min_{x_p, x_t} \lambda\|\mathcal{W}_p\,x_p\|_1 + \lambda\|\mathcal{W}_t\,x_t\|_1 + \frac{1}{2}\|y - x_p - x_t\|_2^2. \tag{3.5}$$

The dichotomy between the synthesis and analysis models [4] is well discussed, for example, in [16]. While in the synthesis formulation signals are modeled as sparse linear combinations of dictionary atoms, the

[3]This last formulation is known in statistics as penalized least square estimation problem.

[4]Note that the two models coincide when the dictionary is an ON basis.

https://doi.org/10.1051/mmnp/20138104 Published online by Cambridge University Press

analysis formulation emphasizes the zeros in the analysis side (rather than the non-zeros), leading to better performance. In particular, one of the major advantages of using the formulation (3.5) rather than (3.4) is that it requires searching lower dimensional vectors rather than longer dimensional representation coefficient vectors.

To further improve the performance, we have also included a total variation regularization term, which is effective at reducing possible ringing artifacts near the edges. This idea was proposed and validated not only in [31], bur also in earlier wavelet papers such as [26]. Thus, we finally have the optimization problem:

$$\hat{x}_p, \hat{x}_t = \arg\min_{x_p, x_t} \lambda \|\mathcal{W}_p\, x_p\|_1 + \lambda \|\mathcal{W}_t\, x_t\|_1 + \gamma\, TV(x_p) + \frac{1}{2}\|y - x_p - x_t\|_2^2, \tag{3.6}$$

where TV is the Total Variation. Once the separate estimates $\hat{x}_p$ and $\hat{x}_t$ are obtained as a solution of (3.6), the final estimator of x is given as $\hat{x} = \hat{x}_p + \hat{x}_t$.

A fast convergent iterative shrinkage algorithm that solves (3.6) may be derived by using Separable Surrogate Functionals (SSF) (see [30] for its derivation) which can be summarized as follows.

We start the algorithm by initializing $k = 1$ and setting $x_p^0 = 0$, $x_t^0 = 0$, $r^0 = y - x_p^0 - x_t^0$, and $\lambda^0 = \frac{1}{2}\left(\|\mathcal{W}_p\, y\|_\infty + \|\mathcal{W}_t\, y\|_\infty\right)$. We repeatedly update the estimate of $\mathbf{x}_p$ and $\mathbf{x}_t$ as

$$\begin{aligned}
\tilde{x}_p^k &= \mathcal{W}_p^\dagger\, \mathcal{T}_{\lambda^k}\left(\frac{1}{c}\,\mathcal{W}_p\, r^{k-1} + \mathcal{W}_p\, x_p^{k-1}\right)\\
x_p^k &= H^\dagger\, \mathcal{T}_{\gamma^k}\left(H\, \tilde{x}_p^k\right)\\
x_t^k &= \mathcal{W}_t^\dagger\, \mathcal{T}_{\lambda^k}\left(\frac{1}{c}\mathcal{W}_t\, r^{k-1} + \mathcal{W}_t\, x_t^{k-1}\right),
\end{aligned}$$

where H is the undecimated Haar wavelet transform. The parameter c is chosen so that it is greater than the maximum eigenvalue of $(\mathcal{W}_p^\dagger\, \mathcal{W}_p + \mathcal{W}_t^\dagger\, \mathcal{W}_t)$. In this particular case, c set to 3 is heuristically found to work well. The residual is updated as

$$r^k = y - x_p^k - x_t^k.$$

and the shrinkage parameter is updated as

$$\lambda^k = \frac{1}{2}\left(\|\mathcal{W}_p\, r^k\|_\infty + \|\mathcal{W}_t\, r^k\|_\infty\right).$$

The Haar shrinkage value γ^k is dependent on the kth iteration since it is set to $3\sigma_{x_p^k}$, where $\sigma_{x_p^k}$ is the standard deviation of the noise estimate of $\tilde{x}_p^k$ found by using a median estimator on the finest scale of the Haar wavelet coefficients. In our numerical tests, the iterations are stopped when λ^k becomes smaller than a chosen threshold. For this procedure, we chose this stopping threshold to be 2.1 times the estimate of the standard deviation of the noisy data.

4. Numerical experiments

To validate our denosing strategy, based on the method described in the previous section, we have built and applied two different types of combined dictionaries, using in one case a combination of our new 3D shearlet transform [29] and local 3D DCT; in another case, a combination of 3D curvelets [2] and local 3D DCT. We denote these dictionary by Shear/DCT and Curv/DCT, respectively.

We have run several numerical experiments for 3D data denoising using as sample data two video sequences that we refer to as *Tempete* and *Oil painting*. All these video sequences, which have been resized to 192^3 can be downloaded from the website `http://www.cipr.rpi.edu`. For a baseline comparison, we tested the performance of the combined-dictionary denoising algorithm against: 3D discrete wavelet transform (DWT) using the "Sym4" filter; 3D curvelet transform (Curv); 3D shearlet transform (Shear); the 3D Laplacian Pyramid (LP). In all these cases, the denoising algorithm is based on the standard soft

https://doi.org/10.1051/mmnp/20138104 Published online by Cambridge University Press

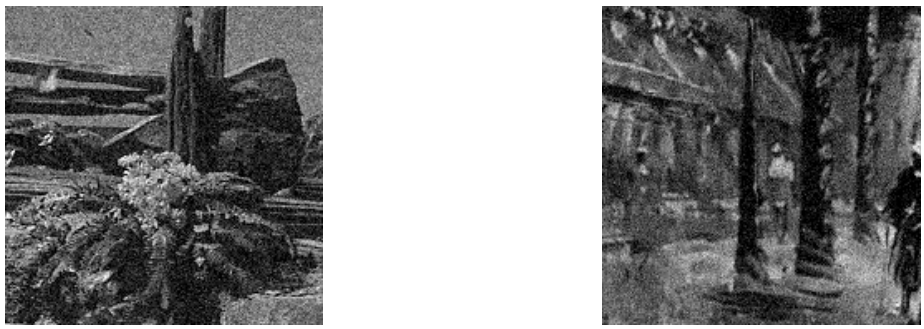


FIGURE 3. Left: the 150th image frame of the *Tempete* video with an additive Gaussian noise whose standard deviation is 20. Right: the 60th image frame of the *Oil painting* video with an additive Gaussian noise whose standard deviation is 20.

thresholding, as described in Section 1 To provide additional baseline comparisons of our combined dictionary techniques, we also expanded the comparisons to include an adaptation of our projection iterative scheme to work with the combinations of the 3D wavelet transform and the local DCT (DWT/DCT) and the 3D Laplacian Pyramid and the local DCT (LP/DCT).

The results of these numerical experiments are given in Tables I and II where we used the *peak signal-to-noise ratio* (PSNR), measured in decibles (dB), given as

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

where $\|\cdot\|_F$ is the Frobenius norm and f is the video array assumed to be of size $N \times N \times N$. The numbers in bold indicate the best performance for the given experiment. Furthermore, we included several figures extracted from the video sequences to illustrate the performance of the various denoising algorithms. Figure 3 shows representative image slices with an additive Gaussian noise incorporated. Illustrations from the *Tempete* and *Oil painting* video are reported in Figures 4 and 5, respectively.

The results reported in this paper show that the new denosing algorithm based on the combination of 3D shearlet and local DCT dictionaries provide the best performance among the methods considered, which are state-of-the-art. As compared with the performance of the 'standard' 3D shearlet (Shear) denosing routine, this result shows the advantage of using a combined rather than a single dictionary, consistently with the general principles which are discussed above (Section 3). Furthermore, our results show that the Shear/DCT routine performs significantly better than similar routines which use combined dictionaries of 3D curvelets and DCT or 3D wavelets and DCT. The improvement with respect to wavelet-based routines is expected, since shearlets provide a sparser dictionary for piecewise smooth data, and their superior performance in denoising applications was already observed in [13, 14, 29]. For the comparison with curvelets, the difference in performance is due mostly to the different strategy used for the implementation of the discrete shearlet transform versus the discrete curvelet transform, where the first transform uses higher redundancy and highly selective directional filter. We refer to [29] for a comparison of 3D shearlets, 3D curvelets, surfacelets and other multiscale transforms in denoising applications using single sparse dictionaries.

In addition to testing the performance of the new denoising routines on video sequences, we have also tested the routine on a 3D scan of a neuronal network dataset that naturally contains noise due to the imaging sensor. In this case, we cannot provide a quantitative measure of the performance, but only

https://doi.org/10.1051/mmnp/20138104 Published online by Cambridge University Press



DWT



Curv



DWT/DCT



Curv/DCT



Shear



Shear/DCT

FIGURE 4. The denoising performance on a representative frame of the video sequence *Tempete* using various denoising routines when the additive Gaussian noise has $\sigma = 20$. Starting from the top left: denoised frame using 3D DWT (PSNR=22.16 dB), 3D Curv (PSNR=22.60 dB), 3D DWT/DCT (PSNR=24.09 dB), 3D Curv/DCT (PSNR=25.29 dB), 3D Shear (PSNR=25.87 dB), and 3D Shear/DCT (27.47 dB).

https://doi.org/10.1051/mmnp/20138104 Published online by Cambridge University Press



LP



DWT/DCT



LP/DCT



Curv/DCT



Shear



Shear/DCT

FIGURE 5. The denoising performance on a representative frame of the video sequence *Oil painting* using various denoising routines when the additive Gaussian noise has $\sigma = 40$. Starting from the top left: denoised frame using 3D LP (PSNR=24.26 dB), 3D DWT/DCT (PSNR=24.89 dB), 3D LP/DCT (PSNR=25.02 dB), 3D Curv/DCT (PSNR=24.67 dB), 3D Shear (PSNR=26.33 dB), and 3D Shear/DCT (27.68 dB).

https://doi.org/10.1051/mmnp/20138104 Published online by Cambridge University Press

Table I: Denoising results (PSNR) using *Tempete video*.

σ	Noisy	DWT	LP	Shear	Curv	Shear/DCT	DWT/DCT	LP/DCT	Curv/DCT
20	22.14	22.16	23.10	25.87	22.60	**27.47**	24.09	24.45	25.29
30	18.62	22.10	22.04	24.63	22.27	**25.61**	22.38	22.61	23.02
40	16.12	20.47	21.30	23.69	22.00	**24.34**	21.40	21.51	21.97

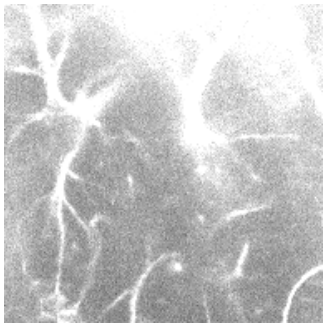
Table II: Denoising results (PSNR) using *Oil painting video*.

σ	Noisy	DWT	LP	Shear	Curv	Shear/DCT	DWT/DCT	LP/DCT	Curv/DCT
20	22.14	26.34	27.01	28.04	27.32	**31.01**	27.74	28.32	27.66
30	18.62	24.81	25.52	27.12	26.86	**29.07**	26.03	26.37	25.94
40	16.12	23.87	24.26	26.33	26.44	**27.68**	24.89	25.02	24.67

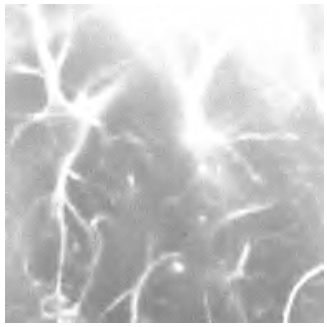
a qualitative assessment based on appearance. Figure 6 displays the results of the various denoising methods for a particular slice of the original 192^3 dataset.

Acknowledgements. DL and PN acknowledge support from NSF grants DMS 1008900 and DMS 1005799. DL also acknowledges support from NHARP grant 003652-0136-2009. The authors thank W. Shain and B. Roysam for providing the biological sample data used to produce Figure 6.

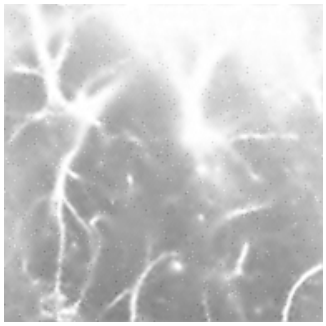
https://doi.org/10.1051/mmnp/20138104 Published online by Cambridge University Press



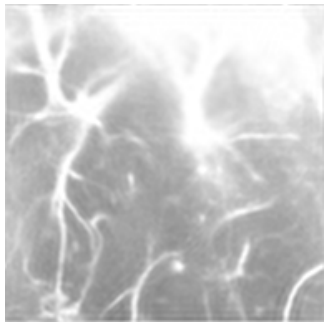
Noisy Data



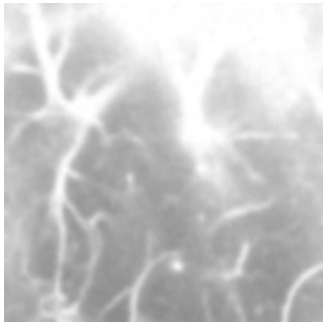
DWT/DCT



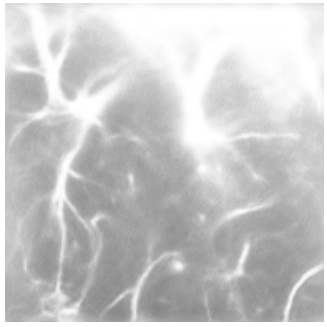
LP/DCT



Curv



Curv/DCT



Shear/DCT

FIGURE 6. The denoising performance on a representative frame of a 3D network of astrocytes using various denoising routines. The routine based on the 3D Shear/DCT combined dictionaries gives the best visual performance.

https://doi.org/10.1051/mmnp/20138104 Published online by Cambridge University Press

References

[1] J. Bobin, J.-L. Starck, M.J. Fadili, Y. Moudden, D.L. Donoho. *Morphological component analysis: an adaptive thresholding strategy*. IEEE Trans. Image Process. 16 (11) (2007), 2675–2681.

[2] E. J. Candès, L. Demanet, D. Donoho, L. Ying. *Fast discrete curvelet transforms*. Multiscale Model. Simul. 5 (2006), 861–899.

[3] E. J. Candès, D. L. Donoho. *Ridgelets: the key to high dimensional intermittency?* Philosophical Transactions of the Royal Society of London A, 357 (1999), 2495–2509.

[4] E. J. Candès, D. L. Donoho. *New tight frames of curvelets and optimal representations of objects with C^2 singularities*. Comm. Pure Appl. Math., 57 (2004), 219–266.

[5] S. S. Chen, D. L. Donoho, M. A. Saunders. *Atomic decomposition by basis pursuit*. SIAM Rev. 43 (1) (2001), 129–159.

[6] I. Daubechies, M. Defrise, C. De Mol. *An iterative thresholding algorithm for linear inverse problems with a sparsity constraint*. Comm. Pure Appl. Math. 57 (2004), 1413–1457.

[7] D. L. Donoho. *Denoising by soft thresholding*. IEEE Trans. Inf. Theory, 41 (3) (1995), 613–627.

[8] D. L. Donoho. *Sparse components of images and optimal atomic decomposition*. Constr. Approx. 17 (2001), 353–382.

[9] D. L. Donoho. *Wedgelets: nearly-minimax estimation of edges*. Annals of Statistics, 27 (1999), 859–897.

[10] D. L. Donoho, I. M. Johnstone. *Ideal spatial adaptation by wavelet shrinkage*. Biometrika, 81 (3) (1994), 425–455.

[11] D. L. Donoho, I. M. Johnstone. *Adapting to unknown smoothness via wavelet shrinkage*. J. Amer. Statist. Assoc. 90 (1995), 1200–1224.

[12] D. L. Donoho, I. M. Johnstone, G. Kerkyacharian, D. Picard. *Wavelet shrinkage. Asymptopia*. J. Roy. Statist. Soc. B, 57 (2) (1995), 301–337.

[13] G. R. Easley, D. Labate, F. Colonna. *Shearlet-based total variation diffusion for denoising*. IEEE Trans. Image Proc. 18 (2) (2009), 260–268.

[14] G. R. Easley, D. Labate, W. Lim. *Sparse directional image representations using the discrete shearlet transform*. Appl. Comput. Harmon. Anal., 25 (1) (2008), 25–46.

[15] M. Elad. *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*. Springer, New York, NY, 2010.

[16] M. Elad, P. Milanfar, R. Rubinstein. *Analysis Versus Synthesis in Signal Priors*. Inverse Problems, 23 (3) (2007), 947–968.

[17] K. Guo, G. Kutyniok, D. Labate. *Sparse Multidimensional Representations using Anisotropic Dilation and Shear Operators,* in: Wavelets and Splines, G. Chen and M. Lai (eds.), Nashboro Press, Nashville, TN (2006), 189–201.

[18] K. Guo, D. Labate. *Optimally Sparse Multidimensional Representation using Shearlets*. SIAM J. Math. Anal.. 9 (2007), 298–318.

[19] K. Guo, D. Labate. *Optimally sparse 3D approximations using shearlet representations*. Electron. Res. Announc. Math. Sci. 17 (2010), 126–138.

[20] K. Guo, D. Labate. *Optimally sparse representations of 3D Data with C^2 surface singularities using Parseval frames of shearlets*. SIAM J Math. Anal. 44 (2012), 851–886.

[21] K. Guo, D. Labate. *The Construction of Smooth Parseval Frames of Shearlets*. Math. Model. Nat. Phenom. 8 (1) (2013), 3255.

[22] X. Huo. *Sparse Image Representation Via Combined Transforms*, Ph.D. Thesis, Stanford University, 1999.

[23] G. Kutyniok. *Clustered sparsity and separation of cartoon and texture*, preprint (2012).

[24] D. Labate, W.-Q Lim, G. Kutyniok, G. Weiss. *Sparse multidimensional representation using shearlets*, in Wavelets XI, edited by M. Papadakis, A. F. Laine, and M. A. Unser, SPIE Proc. 5914 (2005), SPIE, Bellingham, WA, 2005, 254–262.

[25] Y. Lu, M. N. Do. *Multidimensional directional filter banks and surfacelets*, IEEE Trans. Image Process., 16 (4) (2007), 918–931.

[26] F. Malgouyres. *Minimizing the total variation under a general convex constraint for image restoration*, IEEE Trans. Signal Process. 11 (12) (2002), 1450–1456.

[27] S. Mallat. *A Wavelet Tour of Signal Processing.Third Edition: The Sparse Way*, Academic Press, San Diego, CA, 2008.

[28] F. G. Meyer, A. Z. Averbuch, R. Coifman. *Multi-layered image representation: Application to image compression*, IEEE Trans. Image Process. 11(6) (1998), 1072–1080.

[29] P. S. Negi, D. Labate. *3D discrete shearlet transform and video processing*, IEEE Trans. Image Process. 21(6) (2012), 944–2954.

[30] V. M. Patel, G. R. Easley, R. Chellappa, *Component-based restoration of speckled images*, Proceedings 18th IEEE International Conference on Image Processing (ICIP), 2011.

[31] J. L. Starck, M. Elad, D.L. Donoho. *Image decomposition via the combination of sparse representations and a variational approach*, IEEE Trans. Image Process. 14(10) (2005), 1570–1582.

[32] J. L. Starck, F. Murtagh, A. Bijaoui. *Multiresolution support applied to image filtering and restoration*, Graphic. Models Image Process. 57(1995), 420–431.

[33] J. L. Starck, F. Murtagh, J. M. Fadili. *Sparse Image and Signal Processing*, Cambridge University Press, New York, NY, 2010.

[34] A. Woiselle, J. L. Starck, J. M. Fadili. *3-D Data denoising and inpainting with the Low-Redundancy Fast Curvelet Transform*, J. Math. Imaging Vis. 39(2) (2011), 121–139.

https://doi.org/10.1051/mmnp/20138104 Published online by Cambridge University Press