MULTISCALE MODEL. SIMUL.
Vol. 2, No. 4, pp. 554–579

A MULTISCALE IMAGE REPRESENTATION USING HIERARCHICAL (BV, L^2) DECOMPOSITIONS*

EITAN TADMOR†, SUZANNE NEZZAR‡, AND LUMINITA VESE‡

Abstract. We propose a new multiscale image decomposition which offers a hierarchical, adaptive representation for the different features in general images. The starting point is a variational decomposition of an image, $f = u_0 + v_0$, where $[u_0, v_0]$ is the minimizer of a J-functional, $J(f, \lambda_0; X, Y) = \inf_{u+v=f}\{\|u\|_X + \lambda_0\|v\|_Y^p\}$. Such minimizers are standard tools for image manipulations (e.g., denoising, deblurring, compression); see, for example, [M. Mumford and J. Shah, *Proceedings of the IEEE Computer Vision Pattern Recognition Conference*, San Francisco, CA, 1985] and [L. Rudin, S. Osher, and E. Fatemi, *Phys. D*, 60 (1992), pp. 259–268]. Here, u_0 should capture "essential features" of f which are to be separated from the spurious components absorbed by v_0, and λ_0 is a *fixed* threshold which dictates separation of scales. To proceed, we iterate the refinement step $[u_{j+1}, v_{j+1}] = \operatorname{arginf} J(v_j, \lambda_0 2^j)$, leading to the hierarchical decomposition, $f = \sum_{j=0}^{k} u_j + v_k$. We focus our attention on the particular case of $(X, Y) = (BV, L^2)$ decomposition. The resulting hierarchical decomposition, $f \sim \sum_j u_j$, is essentially nonlinear. The questions of convergence, energy decomposition, localization, and adaptivity are discussed. The decomposition is constructed by numerical solution of successive Euler–Lagrange equations. Numerical results illustrate applications of the new decomposition to synthetic and real images. Both greyscale and color images are considered.

Key words. natural images, multiscale expansion, total variation, localization, adaptivity

AMS subject classifications. 26B30, 65C20, 68U10

DOI. 10.1137/030600448

1. Introduction and motivations.

Images could be realized as general L^2 objects, $f \in L^2(\mathbb{R}^2)$, representing the greyscale of the observed image. Likewise, color images are typically realized in terms of vector-valued functions, $\mathbf{f} = (f_1, f_2, f_3) \in L^2(\mathbb{R}^2)^3$, representing the RGB-color scales. In practice, the more noticeable features of images are identified within a proper subclass of all L^2 objects. Most noticeable are the edges of an image, which are known to be well quantified within the smaller subclass of functions of bounded variation (BV), e.g., [4], [3], [5] [6], [11], [16], [17], [23], [24], [25]. The image representation of a real scene often contains other noticeable features, ranging from homogeneous regions to oscillatory patterns of noise or texture. A large class of those images therefore belong to intermediate spaces, lying "between" the larger $L^2(\mathbb{R}^2)$ and the smaller[1] $BV(\mathbb{R}^2)$. Quantifying the precise L^2 subclasses of these different features is still the subject of current research. In this paper we introduce a novel hierarchical, multiscale representation of images. We argue that this new multiscale description is particularly adapted for images lying in such intermediate spaces.

*Received by the editors July 13, 2003; accepted for publication (in revised form) February 25, 2004; published electronically September 2, 2004. This research was supported in part by ONR grant N00014-91-J-1076 (ET) and NSF grant DMS01-07428 (ET, SN), by NSF grant ITR-0113439, and by NIH grant P20MH65166. Part of the research was carried out while S. Nezzar and L. Vese were visiting the Center for Scientific Computation and Mathematical Modeling (CSCAMM) at the University of Maryland, College Park.

http://www.siam.org/journals/mms/2-4/60044.html

†Department of Mathematics, Center of Scientific Computation and Mathematical Modeling (CSCAMM) and Institute for Physical Science and Technology (IPST), University of Maryland, College Park, MD 20742 (tadmor@cscamm.umd.edu).

‡Department of Mathematics, University of California, Los Angeles, CA 90095 (snezzar@math.ucla.edu, lvese@math.ucla.edu).

[1] BV stands for the homogeneous space $BV = \{f \mid \|f\|_{BV} := \sup_{h\neq 0} |h|^{-1}\|f(\cdot + h) - f(\cdot)\|_{L^1} < \infty\}$.

The standard tool for studying intermediate spaces is interpolation, e.g., [8], [9], [13]. To this end, one starts with a pair of given spaces, $Y \subset X$, and forms a scale of intermediate spaces, $(X,Y)_\theta$, $\theta \in [0,1]$, ranging from $(X,Y)_{\theta=0} = X$ to $(X,Y)_{\theta=1} = Y$. The canonical example involves the so-called K-functional

$$K(f,\eta) \equiv K(f,\eta;X,Y) := \inf_{u+v=f} \Big\{ \|v\|_X + \eta \|u\|_Y \Big\}.$$

The space $(X,Y)_\theta$ is dictated by the behavior of $K(\cdot,\eta)$ as $\eta \downarrow 0$—it consists of all f's such that $\{f \mid \eta^{-\theta} K(f,\eta;X,Y) \le \text{Const}\}$. There are many refinements and other variants. For example, refining the L^∞ boundedness of $\eta^{-\theta}K(f,\eta)$ with the requirement $\eta^{-\theta}K(f,\eta;X,Y) \in L^q(d\eta/\eta)$ leads to Lorentz-scale refinement $(X,Y)_{\theta,q}$, depending on a secondary scale q. The K-functional could be replaced by the closely related

$$J_p(f,\eta;X,Y) := \inf_{u+v=f} \Big\{ \|v\|_X^p + \eta \|u\|_Y \Big\},$$

which leads to another variant with a similar scale of intermediate spaces. In the present context of image processing, one seeks the representation in the intermediate spaces between $X = L^2(\Omega)$ and $Y = BV(\Omega)$ defined over two-dimensional domains $\Omega \subset \mathbb{R}^2$ and quantified in terms of the J-functional

$$J(f,\lambda) \equiv J_2(f,\lambda;BV,L^2) := \inf_{u+v=f} \Big\{ \lambda \|v\|_{L^2}^2 + \|u\|_{BV} \Big\}.$$

Because of the reversed order—starting with the smaller $X = BV(\Omega)$ and ending with the larger $L^2(\Omega)$—we shift our focus from small η's to large λ's. The functional $J(f,\lambda)$ measures how well an L^2 object can be approximated by its BV features, $J(f,\lambda) \sim \lambda^\theta$ as $\lambda \uparrow \infty$. The classical argument addresses this question of convergence (or growth) rate of $J(f,\lambda)$ in terms of the smoothness properties of f—an intermediate smoothness between L^2 and BV. In modern theory, however, the roles are reversed: one *defines* the scale of intermediate smoothness spaces such as $(L^2,BV)_\theta$ in terms of the behavior of minimizers such as $K(f,\lambda)$.

The functional $J(f,\lambda)$ was introduced in the present context of image processing by Rudin, Osher, and Fatemi. In their pioneering work, [25], they suggested extracting the main features of contour discontinuities u_λ, which are to be separated from the noisy part v_λ, by realizing the minimizing pair, $[u_\lambda, v_\lambda]$, of $J(f,\lambda)$. In [26], λ is treated as a *fixed* threshold for cutting out the noisy part of f. The cut-off scale λ needs to be predetermined, say, by the known statistical properties of the image under consideration.

The realization of an image f as a minimizing $J(f,\lambda)$-pair, $f = u_\lambda + v_\lambda$, falls within the class of so-called $u+v$ models [19]. There are different perspectives on this question of image processing, using other $u+v$ models. The celebrated Mumford–Shah model [21] is the forerunner of this class. A regularized version of the Mumford–Shah functional was introduced by Ambrosio and Tortorelli [3] and Ambrosio, Fusco, and Pallara [2, section 6]:

$$AT^\varepsilon(f,\lambda) := \inf_{\{w,u,v \mid u+v=f\}} \left\{ \int_\Omega w^2 \Big[|\nabla u|^2 + |v|^2 \Big] dx + \lambda \left[\varepsilon \|\nabla w\|_{L^2}^2 + \frac{\|1-w\|_{L^2}^2}{\varepsilon} \right] \right\}.$$

Letting $\varepsilon \downarrow 0$, then $u = u_\varepsilon$ approaches the Mumford-Shah minimizer while the auxiliary function $1 - w_\varepsilon$ approaches an edge detector for the boundaries enclosing the

objects identified with u_ε. Both $J(\cdot)$ and $AT(\cdot)$ are examples for a larger class of $u+v$ decompositions which are identified as minimizers of appropriate energy functionals. Let us mention another type of $u+v$ decomposition offered by DeVore and Lucier [14]. Again, the noticeable features of an L^2 image f are realized in an intermediate space, this time an intermediate space lying between $X = L^2$ and the (slightly) smaller Besov space $Y = B_1^{1,1}$. The advantage of the scale of spaces spanned by this pair, $(L^2, B_1^{1,1})_\theta$, is that one can efficiently extract and separate scales in terms of a wavelet decomposition of $f = \sum \hat{f}_{jk}\psi_{jk}$. In particular, the wavelet decomposition of such intermediate spaces offers the usual decomposition into the *hierarchy* of dyadic scales. To extract the main features above a fixed scale, one could implement a wavelet shrinkage based on a "greedy" approach of cutting out noisy data by removing wavelets with amplitudes below threshold η, $f \approx \sum_{|\hat{f}_{jk}|\geq\eta} \hat{f}_{jk}\psi_{jk}$. There is no such simple hierarchical description of $(L^2, BV)_\theta$ in terms of truncated wavelet expansion, e.g., [19, section 18, Corollary 1] (but see the recent results of [12]). In particular, no "greedy" algorithm is available with extraction of BV features, similar to what is available with $B_1^{1,1}$. The disadvantage, however, is the failure of $B_1^{1,1}$ to faithfully capture the location of sharp edges.

In this paper we introduce a new multiscale procedure using hierarchical representations, which enables us to capture an intermediate regularity between L^2 and BV. Unlike the one-scale present in $u+v$ decompositions, in our approach λ is not a fixed threshold but varies over a sequence of dyadic scales. Consequently, the representation of an L^2 image is not predetermined but is resolved in terms of layers of intermediate scales. We use (BV, L^2) to symbolize this multilayered representation. The resulting hierarchical representation is outlined in section 2. In section 3 we provide explicit construction for the hierarchical expansion of a few simple objects. In particular, we point out the possibility of making our hierarchical expansion *adaptive*. The hierarchical decomposition of real-life images is simulated in section 4. The hierarchical, multiscale decomposition offered in this paper is not restricted to the J-minimizer of Rudin, Osher, and Fatemi [25]; in section 5 we conclude with extensions based on other minimizers.

2. The hierarchical (BV, L^2) decomposition.

2.1. The hierarchical decomposition. To recover an image from its noisy version f, Rudin, Osher, and Fatemi [25] considered the minimizer $\inf_{u+v=f}\{\lambda\|v\|_{L^2}^2 + \|u\|_{BV}\}$. Here, $\|v\|_{L^2}^2$ is a fidelity term, $\|u\|_{BV}$ is a regularizing term, and $\lambda > 0$ is a weighting parameter, serving as a scaling level to separate the two terms. For $f \in L^2(\Omega)$ the problem admits a unique minimizer (see [10], [1], [27]), which decomposes an $L^2(\Omega)$ image, f, into two distinct components,

$$f = u_\lambda + v_\lambda, \quad [u_\lambda, v_\lambda] = \operatorname*{arginf}_{u+v=f} J(f, \lambda; BV, L^2). \tag{2.1}$$

The BV part, $u = u_\lambda$, captures the main features of f while neglecting the noisy part $v = v_\lambda$. This model is a very effective tool in denoising images while preserving edges. It requires, however, a priori information on the noise scaling λ. Otherwise, if $J(f, \lambda)$ is being implemented with a too small λ, then only a cartoon representation of f is kept in the form of $u_\lambda \in BV$, while small textured patterns or oscillatory details are swept into the residual $v_\lambda := f - u_\lambda$. If λ is kept too large, however, then u_λ remains loaded with too many details, which is close to the original f; not much change has been applied to f, and the compression ratio is small. In some cases,

e.g., [25], [10], the parameter λ can be estimated if some statistical information on the noise is known. In this setup we are limited by the use of the one scale dictated by λ. A multiscale version was introduced by Rudin and Caselles in [24]. We propose a multiscale alternative based on the *hierarchical* image representation of f. We will see that the resulting multiscale decomposition of f enables us to effectively manipulate general images.

Our starting point is an alternative point of view argued by Meyer [19], where the minimization $J(f,\lambda)$ is interpreted as a decomposition, $f = u_\lambda + v_\lambda$, so that u_λ extracts the edges of f while v_λ captures *textures*. Of course, the distinction between these two components is scale dependent—whatever is interpreted as "texture" at a given scale λ consists of significant edges when viewed under a refined scale, say 2λ,

$$v_\lambda = u_{2\lambda} + v_{2\lambda}, \quad [u_{2\lambda}, v_{2\lambda}] = \operatorname*{arginf}_{u+v=v_\lambda} J(v_\lambda, 2\lambda). \tag{2.2}$$

We now have a better two-scale representation of f given by $f \approx u_\lambda + u_{2\lambda}$; texture below scale $1/2\lambda$ remains unresolved in $v_{2\lambda}$. This process (2.2) can continue. Starting with an initial scale $\lambda = \lambda_0$,

$$f = u_0 + v_0, \quad [u_0, v_0] = \operatorname*{arginf}_{u+v=f} J(f, \lambda_0),$$

we proceed with successive application of the dyadic refinement step (2.2),

$$v_j = u_{j+1} + v_{j+1}, \quad [u_{j+1}, v_{j+1}] := \operatorname*{arginf}_{u+v=v_j} J(v_j, \lambda_0 2^{j+1}), \quad j = 0, 1, \ldots, \tag{2.3}$$

producing, after k such steps, the following hierarchical decomposition of f:

$$\begin{aligned} f &= u_0 + v_0 \\ &= u_0 + u_1 + v_1 \\ &= \ldots\ldots \\ &= u_0 + u_1 + \cdots + u_k + v_k. \end{aligned} \tag{2.4}$$

We end up with a new multiscale image decomposition, $f \sim u_0 + u_1 + \cdots + u_k$, with a residual v_k. As k increases, the u_k's resolve edges with increasing scales $\sim \lambda_k := \lambda_0 2^k$. We note in passing that, as usual, coarser and finer decompositions are available, using different ladder of scales, e.g., $\lambda_k = \lambda_0 s^k$, with $1 < s < 2$ (respectively, $s > 2$) leading to finer (respectively, coarser) decompositions of f.

The construction of the hierarchical, multiscale expansion (2.4) is independent of a priori parameters. The partial sum, $\sum_j^k u_j$, provides a multilayered description of f which lies in an intermediate scale of spaces, in between BV and L^2, though the precise regularity may vary, depending on the scales present in f. We use (BV, L^2) to denote a generic intermediate scale space. This multilayered (BV, L^2) expansion, $f \sim \sum_j u_j$, is particularly suitable for image representations. Let us mention applications of multilayered representations to image compression in the context of wavelet expansions that were discussed in [18], [7]. We note that the multilayered representation furnished by (2.4), however, is *essentially nonlinear* in the sense that its dyadic blocks, u_j, depend on the data itself, $u_j = u_j(f)$. These dyadic blocks, $\{u_j\}_{j\geq 0}$, capture different scales of the original image. We turn to quantify the multiscale nature of the hierarchical expansion, $f \sim \sum_j u_j$.

2.2. Convergence of the (BV, L^2) expansion. To quantify the convergence $\sum^k u_j \to f$ as $k \uparrow \infty$, we compare the decomposition of $v_j = u_{j+1} + v_{j+1}$ furnished by the minimizer of $J(v_j, \lambda_{j+1})$, vs. the trivial pair $[0, v_j]$, to find

$$\|u_{j+1}\|_{BV} + \lambda_{j+1}\|v_{j+1}\|_2^2 \le \lambda_{j+1}\|v_j\|_2^2, \quad \lambda_j := \lambda_0 2^j. \tag{2.5}$$

It follows that

$$\begin{aligned}\sum_{j\ge 0}\frac{1}{\lambda_j}\|u_j\|_{BV} &= \frac{1}{\lambda_0}\|u_0\|_{BV} + \sum_{j=0}\frac{1}{\lambda_{j+1}}\|u_{j+1}\|_{BV}\\ &\le \|f\|_2^2 - \|v_0\|_2^2 + \sum_{j=0}\Big[\|v_j\|_2^2 - \|v_{j+1}\|_2^2\Big] \le \|f\|_2^2,\end{aligned} \tag{2.6}$$

in agreement with the fact that the u_j's capture the BV dyadic scales of order $\sim \lambda_j = \lambda_0 2^j$. A more precise (BV, L^2) hierarchical statement is provided in the following.

THEOREM 2.1. *Consider $f \in L^2$. Then f admits the following hierarchical decomposition:*[2]

$$f = \sum_{j=0}^{\infty} u_j, \quad \Big\|f - \sum_{j=0}^{k} u_j\Big\|_{W^{-1,\infty}} = \frac{1}{\lambda_0 2^{k+1}}, \tag{2.7}$$

and the following "energy" estimate holds:

$$\sum_{j=0}^{\infty}\Big[\frac{1}{\lambda_j}\|u_j\|_{BV} + \|u_j\|_2^2\Big] \le \|f\|_2^2, \quad \lambda_j := \lambda_0 2^j. \tag{2.8}$$

Proof. We begin by quoting the following characterization of the $J(f, \lambda)$ minimizer [19, Theorem 3], depending on the oscillatory part of f which is measured by its $W^{-1,\infty}$ norm. Namely, if $\|f\|_{W^{-1,\infty}} < 1/2\lambda$, then $[u_\lambda, v_\lambda] = [0, f]$; otherwise,

$$\|v_\lambda\|_{W^{-1,\infty}} = \frac{1}{2\lambda}, \quad (u_\lambda, v_\lambda) := \int u_\lambda(x)v_\lambda(x)dx = \frac{1}{2\lambda}\|u_\lambda\|_{BV}. \tag{2.9}$$

We observe that according to (2.9), the minimizer $[u_\lambda, v_\lambda]$ becomes an *extremal pair* by placing an equality in the duality statement $\int g(x)h(x)dx \le \|g\|_{W^{-1,\infty}}\|h\|_{BV}$ (the latter follows by a density argument, starting from the usual duality between $W^{-1,\infty}$ and $W^{1,1}$).

The first statement (2.7) then follows from the basic hierarchical expansion, $f = \sum_0^k u_j + v_k$, while noting that $\|v_k\|_{W^{-1,\infty}} = 1/2\lambda_k$. For the second statement, (2.8), we begin by squaring the basic refinement step, $u_{j+1} + v_{j+1} = v_j$,

$$\|v_{j+1}\|_2^2 + \|u_{j+1}\|_2^2 + 2(u_{j+1}, v_{j+1}) = \|v_j\|_2^2, \quad j = -1, 0, 1, \dots . \tag{2.10}$$

Observe that the last equality holds for $j = -1$ with v_{-1} interpreted as $v_{-1} := f$. We recall that (u_{j+1}, v_{j+1}) is a minimizing pair for $J(v_j, \lambda_{j+1})$, and hence, by (2.9),

$$2(u_{j+1}, v_{j+1}) = \frac{1}{\lambda_{j+1}}\|u_{j+1}\|_{BV},$$

[2] We employ the usual notation, $\|f\|_{W^{-1,\infty}} := \sup_g[\int f(x)g(x)/\|g\|_{W^{1,1}}]$, $\|g\|_{W^{1,1}} := \|\nabla g\|_{L^1}$.

yielding $\frac{1}{\lambda_{j+1}}\|u_{j+1}\|_{BV} + \|u_{j+1}\|_2^2 = \|v_j\|_2^2 - \|v_{j+1}\|_2^2$ (which is a precise refinement of (2.5)). We sum up, obtaining

$$\begin{aligned}\sum_{j=0}^{k}\left[\frac{1}{\lambda_j}\|u_j\|_{BV} + \|u_j\|_2^2\right] &= \sum_{j=-1}^{k-1}\left[\frac{1}{\lambda_{j+1}}\|u_{j+1}\|_{BV} + \|u_{j+1}\|_2^2\right] \\ &= \|v_{-1}\|_2^2 - \|v_k\|_2^2 = \|f\|_2^2 - \|v_k\|_2^2. \qquad \square \end{aligned} \tag{2.11}$$

We note that the statement (2.7) is limited to weak convergence of the hierarchical decomposition, $f \sim \sum_j u_j$. Yet, measured in this weak $W^{-1,\infty}$ topology, the geometric convergence rate is *universal*, independent of $f \in L^2$. This universality is due to the nonlinearity of the hierarchical decomposition (2.7). To convert this statement into a strong convergence, we seek an equality in the energy inequality (2.8). According to (2.11), equality holds iff we have strong L^2 convergence, $\|f - \sum^k u_j\|_2 = \|v_k\|_2 \to 0$. The situation is reminiscent of the passage, in the linear setup, from the Bessel-energy inequality into the Parseval equality. Since the present setup is nonlinear, the linear sense of completeness of $\{u_j(f)\}_{j\geq 0}$ does not apply. Instead, we show energy equality and strong L^2 convergence by adding minimal amount of smoothness. We begin with the following.

THEOREM 2.2. *Consider $f \in BV$. Then the (BV, L^2) hierarchical decomposition of f, $f = \sum_{j=0}^{\infty} u_j$, converges strongly in L^2, and the energy of f is given by*

$$\sum_{j=0}^{\infty}\left[\frac{1}{\lambda_j}\|u_j\|_{BV} + \|u_j\|_2^2\right] = \|f\|_2^2. \tag{2.12}$$

Proof. Recall that v_k denotes the "texture" at scale λ_k and that according to (2.11), we have to show the strong convergence $\|v_k\|_2 \to 0$. Our starting point is the decomposition $v_{2k} = -\sum_{j=k+1}^{2k} u_j + v_k$. Multiplication against v_{2k} yields

$$\|v_{2k}\|_2^2 = -\left(v_{2k}, \sum_{j=k+1}^{2k} u_j\right) + (v_{2k}, v_k) =: I + II. \tag{2.13}$$

Recall that the $W^{-1,\infty}$ norm of v_{2k} is given by $1/2\lambda_{2k}$ so that $|(v_{2k}, h)| \leq \|h\|_{BV}/2\lambda_{2k}$. We find that the first term on the right-hand side of (2.13), $I = -(v_{2k}, \sum_{j=k+1}^{2k} u_j)$, does not exceed

$$|I| \leq \frac{1}{2\lambda_{2k}} \sum_{j=k+1}^{2k} \|u_j\|_{BV} \leq \sum_{j=k+1}^{2k} \frac{1}{2\lambda_j}\|u_j\|_{BV},$$

and hence it decays to zero for $k \uparrow \infty$ as a Cauchy subsequence of the bounded series $\sum \frac{1}{\lambda_j}\|u_j\|_{BV} \leq \|f\|_2^2$; see (2.6). It remains to treat the second term, (v_{2k}, v_k). To this end we note that the BV norm of v_k does not grow faster than 2^k; indeed, since $v_k = f - \sum_{j=0}^{k} u_j$, we have the upper bound

$$\|v_k\|_{BV} \leq \|f\|_{BV} + \sum_{j=0}^{k} \|u_j\|_{BV} \leq \|f\|_{BV} + \lambda_k \sum_{j=0}^{k} \frac{1}{\lambda_j}\|u_j\|_{BV} \leq \|f\|_{BV} + \lambda_k \|f\|_2^2. \tag{2.14}$$

We conclude that the second term on the right-hand side of (2.13), $II = (v_{2k}, h)$, with $h = v_k$ vanishes as $k \uparrow \infty$:

$$|II| \le \frac{1}{2\lambda_{2k}}\|v_k\|_{BV} \le \frac{1}{2\lambda_{2k}}\Big[\|f\|_{BV} + \lambda_k\|f\|_2^2\Big] \downarrow 0. \qquad \square$$

It is clear that the last result can be extended for f's beyond the BV space. To this end, let us revisit the estimate (2.13). The first one, I, vanishes for arbitrary $f \in L^2$, and we need to treat only the second term on its right-hand side, $II = (v_{2k}, v_k)$, which is upper bounded by

$$|(v_{2k}, v_k)| = \left|(v_{2k}, f) - \sum_{j=0}^{k}(v_{2k}, u_j)\right| \le |(v_{2k}, f)| + \frac{\lambda_k}{2\lambda_{2k}}\|f\|_2^2.$$

Thus, the energy statement (2.12) holds iff the moments $|(v_k, f)| \to 0$.

To satisfy this vanishing moments condition, let us assume that f belongs to the interpolation space $X_\theta := (L^2, BV)_\theta$, $\theta > 0$. Characterization of this scale of space can be found in [12]. Let $X_{-\theta}$ denote the dual space, the collection of all f's such that $\|f\|_{X_{-\theta}} := \sup_g \int f(x)g(x)/\|g\|_{X_\theta} < \infty$. We recall that v_{2k} is L^2 bounded, $\|v_{2k}\|_2 \le \|f\|_2$, while its $W^{-1,\infty}$ size is given by $\|v_{2k}\|_{W^{-1,\infty}} = 1/2\lambda_{2k}$. By a convexity argument of Riesz we find

$$\|v_{2k}\|_{X_{-\theta}} \le \text{Const}\|v_{2k}\|_2^{1-\theta}\|v_{2k}\|_{W^{-1,\infty}}^{\theta} \le \text{Const}\|f\|_2^{1-\theta}2^{-2k\theta}.$$

We conclude that

$$|(v_{2k}, f)| \le \|v_{2k}\|_{X_{-\theta}}\|f\|_{X_\theta} \le \text{Const}\|f\|_2^{1-\theta}\|f\|_{X_{-\theta}} \cdot 2^{-2k\theta},$$

which in turn implies strong L_2 convergence of texture terms, $\|v_k\|_2 \to 0$, and the desired energy statement (2.12) follows. We summarize by stating the following.

COROLLARY 2.3. *Consider $f \in (L^2, BV)_\theta$, $\theta > 0$. Then the (BV, L^2) hierarchical decomposition of f, $f = \sum_{j=0}^{\infty} u_j$, converges strongly in L^2, and the energy of f is given by*

$$\|f\|_2^2 = \sum_{j=0}^{\infty}(f, u_j) = \sum_{j=0}^{\infty}\left[\frac{1}{\lambda_j}\|u_j\|_{BV} + \|u_j\|_2^2\right]. \tag{2.15}$$

Other extensions along these lines are possible. The flavor is the same; namely, a minimal amount of smoothness beyond the L^2 bound will guarantee strong convergence. The question of strong convergence for $f \in L^2$, corresponding to $\theta = 0$, remains open.

Finally, let us note that the decomposition of energy stated in (2.15) lies entirely with the BV scales. Specifically, we have

$$\sum_{j=0}^{\infty}\frac{1}{\lambda_j}\|u_j\|_{BV} < \|f\|_2^2 < \frac{3}{2}\sum_{j=0}^{\infty}\frac{1}{\lambda_j}\|u_j\|_{BV}. \tag{2.16}$$

We need to address only the upper bound on the right-hand side. By duality, $\|u_j\|_2^2 \le \|u_j\|_{W^{-1,\infty}}\|u_j\|_{BV}$. But $u_j = v_{j-1} - v_j$ implies that the $W^{-1,\infty}$ size of u_j does not exceed

$$\|u_j\|_{W^{-1,\infty}} \le \frac{1}{2\lambda_{j-1}} - \frac{1}{2\lambda_j} = \frac{1}{2\lambda_j}.$$

The bound could be viewed as the dual estimate (2.14) for the growth of v_k. We conclude that $\|u_j\|_{BV}/\lambda_j + \|u_j\|_2^2 \le 3\|u_j\|_{BV}/2\lambda_j$, and (2.16) follows.

2.3. Initialization. How should one choose the initial scale λ_0? Since the optimal $J(f,\lambda)$ decomposition of f for which $\|f\|_{W^{-1,\infty}} \leq 1/2\lambda$ is given by the trivial pair, $[u,v]=[0,f]$, the initial scale, λ_0, should capture smallest oscillatory scale in f, furnished by

$$\frac{1}{2\lambda_0} \leq \|f\|_{W^{-1,\infty}} \leq \frac{1}{\lambda_0}. \tag{2.17}$$

In general, we may not have a priori information on the size of $\|f\|_{W^{-1,\infty}}$. If the initial choice of λ_0 proved to be too small, then the minimizer will remain the same fully textured pair $[u_k,v_k]=[0,f]$ as k increases until a dyadic multiple of λ_0 is large enough so that (2.17) holds. If, on the other hand, the initial λ_0 is chosen too large, we can proceed by a refinement procedure which aims to capture a hierarchical representation of the missing larger scales. We set

$$v_j = u_{j-1} + v_{j-1}, \quad [u_{j-1}, v_{j-1}] := \operatorname*{arginf}_{u+v=v_j} J(v_j, \lambda_{j-1}), \quad j = 0, -1, \ldots . \tag{2.18}$$

We compare the decomposition of $v_j = u_{j-1} + v_{j-1}$ furnished by the optimal pair $[u_{j-1}, v_{j-1}]$ minimizing $J(v_j, \lambda_{j-1})$, vs. the trivial pair $[0, v_j]$, to find

$$\|u_{j-1}\|_{BV} + \lambda_{j-1}\|v_{j-1}\|_2^2 \leq \lambda_{j-1}\|v_j\|_2^2, \quad j = 0, -1, \ldots . \tag{2.19}$$

It follows that

$$\sum_{j\leq 0} \frac{1}{\lambda_{j-1}} \|u_{j-1}\|_{BV} \leq \sum_{j\leq 0} \Big[\|v_j\|_2^2 - \|v_{j-1}\|_2^2 \Big] \leq \|v_0\|_2^2 \leq \lambda_0 \|f\|_2^2, \tag{2.20}$$

which shows the geometric convergence of the dyadic scales captured by the u_j's, for $j = 0, -1, \ldots$

$$\|u_j\|_{BV} \leq \lambda_j \|f\|_2^2, \quad j = 0, -1, \ldots, -k_0. \tag{2.21}$$

As j decreases, the expansion is running through smaller scales, $\lambda_j = \lambda_0 2^j$, until we exhaust the oscillatory part of f by satisfying $\lambda_0 2^{-k_0} \|f\|_{W^{-1,\infty}} \leq 1$. We end up with the hierarchical decomposition

$$\begin{aligned} v_0 &= u_{-1} + v_{-1} \\ &= u_{-1} + u_{-2} + v_{-2} \\ &= \ldots\ldots \\ &= u_{-1} + u_{-2} + \cdots + u_{-k_0}. \end{aligned} \tag{2.22}$$

The multiscale (BV, L^2) expansion now reads

$$f = \sum_{j=-k_0}^{\infty} u_j,$$

with equality understood in the weak $W^{-1,\infty}$ (respectively, L^2) sense for general f's in L^2 (respectively, BV).

2.4. A dual (BV, L^2) expansion. The hierarchical decomposition discussed so far was based on a dyadic refinement of "texture" in terms of "edges." The procedure can be transposed. Let $[u_\lambda, v_\lambda]$ be the minimizer of $J(f, \lambda)$, and consider the resolution of the main features on scale λ, this time in terms of a refined scale of "texture," namely, $u_\lambda = u_{2\lambda} + v_{2\lambda}$. This leads to a dual (BV, L^2) hierarchical expansion of the form

$$f = \sum_{j=0}^{k} v_j + u_k, \quad [u_{j+1}, v_{j+1}] := \operatorname*{arginf}_{u+v=u_j} J(u_j, \lambda_{j+1}). \tag{2.23}$$

The quantitative behavior of this expansion can be worked out as before. To sketch the details, we first compare the optimal pair, $[u_{j+1}, v_{j+1}]$ vs. the trivial one, $[u_j, 0]$, yielding $\lambda_{j+1}\|v_{j+1}\|_2^2 + \|u_{j+1}\|_{BV} \le \|u_j\|_{BV}$; hence

$$\lambda_j \|v_j\|_2^2 + \|u_j\|_{BV} \le \|u_1\|_{BV} \le \lambda_0 \|f\|_2^2.$$

It follows that $\|v_j\|_2 \le \sqrt{\lambda_0/\lambda_j}\|f\|_2$. This dictates an initial scale λ_0 for the dual expansion (2.23), $\lambda_0 \sim 1/\|f\|_2$; otherwise, the expansion is truncated at smaller scales where $v_j = 0$ since $\|v_j\|_{W^{-1,\infty}} < 1/2\lambda_j$. For decomposition of the energy, $\|f\|_2^2$, we square $u_j = u_{j+1} + v_{j+1}$ to find (with $f := u_{-1}$)

$$\|u_j\|_2^2 = \|u_{j+1}\|_2^2 + \frac{1}{\lambda_{j+1}}\|u_{j+1}\|_{BV} + \|v_{j+1}\|_2^2, \quad j = -1, 0, 1 \dots .$$

The telescoping sum then yields

$$\|f\|_2^2 - \|u_k\|_2^2 = \sum_{j=0}^{k} \left[\frac{1}{\lambda_j}\|u_j\|_{BV} + \|v_j\|_2^2 \right], \tag{2.24}$$

and strong convergence follows, provided $\|u_k\| \to 0$.[3] To put this into perspective, we recall the classical K-interpolation spaces, $(L^2, BV)_{\theta,q}$, for $\theta, 1/q \in [0,1]$, which consist of all f's such that $\sum_j (2^{j\theta} K(f, 2^{-j}; L^2, BV))^q < \infty$. Likewise, we define the intermediate scale of spaces $J(BV, L^2)_{\theta,q}$ associated with the finite sum $\sum_j (\lambda_j^{-\theta} J(u_j, \lambda_j; BV, L^2))^q < \infty$. The summability on the right-hand side of (2.24) corresponds to the case $(\theta, q) = (1, 1)$.

3. Examples of (BV, L^2) expansions.

3.1. Hierarchical decomposition over $\mathbb{R}^2$. We begin with the simple example of the characteristic function of a disc, $f(x) = \alpha\chi_{B_R}(x)$, $x \in \mathbb{R}^2$. To illustrate the hierarchical expansion (2.22) in this case, we refer to the optimal $J(f, \lambda)$ decomposition given in [19, Lemma 6],

$$u_\lambda = \left(\alpha - \frac{1}{\lambda R}\right)_+ \chi_{B_R}, \; v_\lambda = f - u_\lambda, \quad [u_\lambda, v_\lambda] = \operatorname*{arginf}_{u+v=\alpha\chi_{B_R}} J(\alpha\chi_{B_R}, \lambda). \tag{3.1}$$

The point here is that already for a simple BV function without any noise such as $f = \alpha\chi_{B_R}$, its J-minimizer at any level $\lambda \ge 1/R\alpha$ contains both a BV part $u_\lambda = (\alpha - \frac{1}{\lambda R})\chi_{B_R}$ and a texture part given by the residual $v_\lambda = \frac{1}{\lambda R}\chi_{B_R}$. Indeed, as

[3] Equating the $W^{-1,\infty}$ norms of (2.23) restricts λ_0 such that $\|f\|_{W^{-1,\infty}} \le \sum_j 1/2\lambda_j = 1/\lambda_0$.

pointed out by Meyer [19, section 1.14], the decomposition of Rudin et al. does *not* keep original BV images; for example, it cannot recover the characteristic function of any Lipschitz domain. Instead, the BV portion of an image is extracted at a given λ scale. Underresolved features are considered "textures." This undesired phenomenon is due to restriction to one scale and should be contrasted with the multiscale representation constructed below. With $\|\alpha\chi_{B_R}\|_{W^{-1,\infty}} = \alpha R/2$, we set as our initial scale

$$\frac{1}{\alpha R} \le \lambda_0 \le \frac{2}{\alpha R},$$

leading to the decomposition at level λ_0,

$$f = u_0 + v_0, \quad u_0 = \left(\alpha - \frac{1}{\lambda_0 R}\right)\chi_{B_R}, \quad v_0 = \frac{1}{\lambda_0 R}\chi_{B_R}.$$

Successive decompositions, $\operatorname{arginf}_{u+v=v_{j-1}} J(v_{j-1}, \lambda_j)$, with $\lambda_j := \lambda_0 2^j$ yield the corresponding minimizers, $[u_j, v_j]$,

$$u_j = \left(\frac{1}{\lambda_{j-1}R} - \frac{1}{\lambda_j R}\right)\chi_{B_R}, \quad v_j = \frac{1}{\lambda_j R}\chi_{B_R}, \quad j = 1, 2 \dots .$$

We end up with hierarchical decomposition, $\alpha\chi_{B_R} = \sum_{j=0}^{k} u_j + v_k$,

$$\alpha\chi_{B_R} \sim u_0 + \sum_{j=1}^{k}\left(\frac{1}{\lambda_{j-1}R} - \frac{1}{\lambda_j R}\right)\chi_{B_R} = \left(\alpha - \frac{1}{\lambda_0 2^j R}\right)\chi_{B_R}. \tag{3.2}$$

The error encountered after k steps is given by $v_k = \frac{1}{\lambda_0 2^k R}\chi_{B_R}$. The convergence is geometric; in this case $\|v_k\|_2 \sim 2^{-k}$.

3.2. Hierarchical decomposition over bounded domains. Consider the characteristic function $f = \alpha\chi_{B_R}$ defined over a bounded domain $\Omega \supset B_R$,

$$f(x) = \alpha\chi_{B_R}(x) := \begin{cases} 1, & |x| \le R, \\ 0, & x \in \Omega \setminus B_R. \end{cases}$$

Then the corresponding minimizer $[u_\lambda, v_\lambda]$ of $J(f, \lambda)$ is given by (here and below, $|\cdot|$ denotes the area of a two-dimensional set)

$$u_\lambda = \left(\alpha - \frac{1}{\lambda R}\right)_+ \chi_{B_R} + \frac{1}{\lambda R}\frac{|B_R|}{|\Omega \setminus B_R|}\chi_{\Omega\setminus B_R},$$
$$v_\lambda := f - u_\lambda = \frac{1}{\lambda R}\chi_{B_R} - \frac{1}{\lambda R}\frac{|B_R|}{|\Omega \setminus B_R|}\chi_{\Omega\setminus B_R}, \quad \lambda > 1/\alpha R.$$

Observe that the natural boundary condition, $\partial u_\lambda/\partial n_{|\partial\Omega} = 0$ (see (4.2) below), requires v_λ to satisfy the consistency condition,

$$\int_\Omega v_\lambda dx = -\frac{1}{2\lambda}\int_{\partial\Omega}\frac{\frac{\partial u_\lambda}{\partial n}}{|\nabla u_\lambda|}dS = 0, \tag{3.3}$$

which in turn dictates the unique, nonzero constant value of v_λ outside the ball B_R. The general hierarchical step then reads

$$u_j = \left(\frac{1}{\lambda_{j-1}R} - \frac{1}{\lambda_j R}\right)\chi_{B_R} + \left(\frac{1}{\lambda_j R} - \frac{1}{\lambda_{j-1}R}\right)\frac{|B_R|}{|\Omega \setminus B_R|}\chi_{\Omega\setminus B_R}, \tag{3.4}$$
$$v_j = \frac{1}{\lambda_j R}\chi_{B_R} - \frac{1}{\lambda_j R}\frac{|B_R|}{|\Omega \setminus B_R|}\chi_{\Omega\setminus B_R}, \qquad j = 1, 2, \dots .$$

We conclude with the (BV, L^2) hierarchical expansion

$$\alpha\chi_{B_R}(x) \sim \sum_{j=0}^{k} u_j = \left(\alpha - \frac{1}{\lambda_k R}\right)\chi_{B_R} + \frac{1}{\lambda_k R}\frac{|B_R|}{|\Omega \setminus B_R|}\chi_{\Omega\setminus B_R} \tag{3.5}$$

with a geometrically vanishing error, $\|v_k\|_2 \sim 1/\lambda_k$.

Remark. In the last two examples we find that the k-step hierarchical decompositions, $f \sim \sum_{j=1}^{k} u_j$, coincide with the one step (BV, L^2) decomposition at the λ_k-scale, i.e., $\sum_{j=1}^{k} u_j = U_k$, where $[U_k, V_k] = \operatorname{arginf}_{u+v=f} J(f, \lambda_k)$. The situation is rather special for the two examples of simple characteristic functions. With general images, however, the two decompositions are different, as shown by numerical experiments carried out in [22], yet they are close to each other in the sense that

$$\left\|\sum_{j=1}^{k} u_j - U_k\right\|_{W^{-1,\infty}} = \|V_k - v_k\|_{W^{-1,\infty}} \le \frac{1}{\lambda_k}.$$

In fact, the distinction between the two decompositions is more fundamental than their sheer size. The (BV, L^2) decomposition introduced in [25] by Rudin, Osher, and Fatemi as a denoising process involves one step to remove the layer of noise V_k from the main signal U_k. The approach taken in this work is different: we seek the decomposition into several layers of signals and textures. Our starting point is that whatever is interpreted as "texture" in a given scale consists of significant features on a finer scale. This is consistent with the one layer $u+v$ modeling of texture suggested by Meyer [19] and the remark made in [20, pp. 93–94] that modeling noise in a given scale follows the same statistics of the cartoon, only scaled down. The one-step signal U_k is therefore replaced by a multilayer of scales, $\sum_{j=1}^{k} u_j$, which are necessary to capture different features at the different scales. Although $\sum_{j=1}^{k} u_j$ is close to U_k, the interpretation is different. Moreover, thanks to the energy estimate (2.15), one can manipulate the hierarchical decomposition, $f \sim \sum_{j=1}^{k} u_j$. As an example, we consider below adaptive domain decompositions where a different number of hierarchical layers apply to different parts of the image.

3.3. Localization and adaptivity. The last example shows that the $J(f, \lambda)$ minimizer need *not* be local in the sense that the support of u_λ could spread well beyond the support of f. Nevertheless, the example discussed in section 3.2 shows that the corresponding hierarchical expansion compensates for localization as the amplitude of $f - \sum^{k} u_j$ decays outside $\operatorname{supp}(f)$. In this context we raise a more general question. Consider an image f as a direct sum, $f = g + h$, where g and h have disjoint supports, $\operatorname{supp}(g) \cap \operatorname{supp}(h) = \emptyset$, and assume g and h admit the (BV, L^2) hierarchical decompositions $g \sim \sum g_j$ and $h \sim \sum h_j$. What can be said about the sum $\sum(g_j + h_j)$ as a hierarchical expansion of f? Clearly, $\|f - \sum^{k}(g_j + h_j)\|_{W^{-1,\infty}} \le 1/\lambda_k$;

the main issue is to quantify strong convergence and, in particular, the behavior of $\mathrm{supp}(g_j)$ and $\mathrm{supp}(h_j)$ relative to $\mathrm{supp}(f)$. The special case $(g,h)=(f,0)$ corresponds to the question of localization. The general case is related to the issue of *adaptivity* of the hierarchical (BV,L^2) expansion.

The following example demonstrates this point. Following Meyer [19, section 1.14], we set $f=\chi_A(x)+p(2^Nx)\chi_B(x)$ with nonintersecting A and B. The function p is assumed 2π-periodic, so that $h\equiv h_N=p(2^Nx)\chi_B(x)$ represents the "noisy part" of f with increasing N, while $g=\chi_A(x)$ represents the "essential feature" in f. If $2^N\gg\lambda$, then the "u-component" of the $J(h,\lambda)$ minimizer fails to separate the essential part of h, since $\|h\|_{W^{-1,\infty}}\sim 2^{-N}<1/2\lambda$. Thus, we need at least $k\sim N$ terms before the hierarchical expansion, $h\sim\sum^k h_j$, would remove the noisy part. On the other hand, the expansion of g is independent of N for $\|g-\sum^k g_j\|\sim 1/\lambda_k$, and we are led to the question of how $\sum^k(g_j+h_j)$ compares with the direct expansion, $f\sim\sum^k u_j$. One way to circumvent the possible global effect of localized oscillations (such as those represented by $h_N(x)$) would be to introduce a localized hierarchical expansion which is adapted to the behavior of f in each subdomain. An adaptive domain decomposition procedure along these lines is discussed in section 4.4.

4. Numerical discretization and experimental results. In this section, we provide the details of the numerical algorithm we used for the construction of our hierarchical decompositions. In each step, we use finite-difference discretization of the Euler–Lagrange equations associated with the $J(v_j,\lambda_{j+1})$ minimization to resolve the next term, u_{j+1}, in the hierarchical decomposition. Numerical results of hierarchical decompositions applied to both synthetic and real images are presented.

4.1. Euler–Lagrange equations. To construct the hierarchical representation of f, we seek the characterization for the minimizer of $J(f,\lambda)$ in terms of the corresponding Euler–Lagrange equation; see, e.g., [4], [10] and the references therein:

$$u_\lambda-\frac{1}{2\lambda}\mathrm{div}\left(\frac{\nabla u_\lambda}{|\nabla u_\lambda|}\right)=f. \tag{4.1}$$

When restricted to a bounded domain Ω, the Euler–Lagrange equations are augmented by the Newman boundary condition

$$\frac{\partial u_\lambda}{\partial n}_{|\partial\Omega}=0. \tag{4.2}$$

This leads to the hierarchical expansion, $f\sim\sum_{j=0}^k u_j$, where the u_j's are constructed as (approximate) solutions of the recursive relation governed by the elliptic PDE,

$$u_{j+1}-\frac{1}{2\lambda_{j+1}}\mathrm{div}\left(\frac{\nabla u_{j+1}}{|\nabla u_{j+1}|}\right)=-\frac{1}{2\lambda_j}\mathrm{div}\left(\frac{\nabla u_j}{|\nabla u_j|}\right). \tag{4.3}$$

4.2. Numerical discretization of Euler–Lagrange equations. We begin by regularization. To remove the singularity when $|\nabla u_\lambda|=0$, we replace $J(f,\lambda)$ by

$$J^\varepsilon(f,\lambda):=\inf_{u+v=f}\left\{\lambda\|v\|^2_{L^2(\Omega)}+\int_\Omega\sqrt{\varepsilon^2+|\nabla u|^2}dxdy\right\}.$$

At each step of our hierarchical scheme, we find the minimizer, $u_\lambda\equiv u_{\lambda,\varepsilon}$, of the regularized functional associated with J^ε. The corresponding Euler–Lagrange equations

read

$$u_\lambda - \frac{1}{2\lambda}\mathrm{div}\left(\frac{\nabla u_\lambda}{\sqrt{\varepsilon^2+|\nabla u_\lambda|^2}}\right) = f \ \text{ in } \Omega, \tag{4.4}$$

$$\frac{\partial u_\lambda}{\partial n} = 0 \ \text{ on } \partial\Omega. \tag{4.5}$$

We cover the domain Ω with a computational grid ($x_i := ih$, $y_j := jh$), with cell size h. We let $D_+ = D_+(h)$, $D_- = D_-(h)$, and $D_0 := (D_+ + D_-)/2$ denote the usual forward, backward, and centered divided difference, e.g., differencing in the x- and y-directions $(D_{\pm x}u)_{i,j} = \pm(u_{i\pm1,j} - u_{i,j})/h, (D_{0y}u)_{i,j} = (u_{i,j+1} - u_{i,j-1})/2h$.

The resulting nonlinear elliptic PDE (4.4) is discretized in a straightforward manner (see [25], [27], and [6]):

$$\begin{aligned}
u_{i,j} &= f_{i,j} + \frac{1}{2\lambda}D_{-x}\left[\frac{1}{\sqrt{\varepsilon^2+(D_{+x}u_{i,j})^2+(D_{0y}u_{i,j})^2}}D_{+x}u_{i,j}\right] \\
&\quad + \frac{1}{2\lambda}D_{-y}\left[\frac{1}{\sqrt{\varepsilon^2+(D_{0x}u_{i,j})^2+(D_{+y}u_{i,j})^2}}D_{+y}u_{i,j}\right] \\
&= f_{i,j} + \frac{1}{2\lambda h^2}\left[\frac{u_{i+1,j}-u_{i,j}}{\sqrt{\varepsilon^2+(D_{+x}u_{i,j})^2+(D_{0y}u_{i,j})^2}} - \frac{u_{i,j}-u_{i-1,j}}{\sqrt{\varepsilon^2+(D_{-x}u_{i,j})^2+(D_{0y}u_{i-1,j})^2}}\right] \\
&\quad + \frac{1}{2\lambda h^2}\left[\frac{u_{i,j+1}-u_{i,j}}{\sqrt{\varepsilon^2+(D_{0x}u_{i,j})^2+(D_{+y}u_{i,j})^2}} - \frac{u_{i,j}-u_{i,j-1}}{\sqrt{\varepsilon^2+(D_{0x}u_{i,j-1})^2+(D_{-y}u_{i,j})^2}}\right].
\end{aligned} \tag{4.6}$$

4.3. The hierarchical (BV, L^2) decomposition scheme for greyscale images. One can use the fixed point Jacobi or Gauss–Seidel iterative methods for solving the discrete regularized Euler–Lagrange equations (4.6). For the former we have

$$\begin{aligned}
u^{n+1}_{i,j} &= f_{i,j} + \frac{1}{2\lambda h^2}\left[\frac{u^n_{i+1,j}-u^{n+1}_{i,j}}{\sqrt{\varepsilon^2+(D_{+x}u^n_{i,j})^2+(D_{0y}u^n_{i,j})^2}} - \frac{u^{n+1}_{i,j}-u^n_{i-1,j}}{\sqrt{\varepsilon^2+(D_{-x}u^n_{i,j})^2+(D_{0y}u^n_{i-1,j})^2}}\right] \\
&\quad + \frac{1}{2\lambda h^2}\left[\frac{u^n_{i,j+1}-u^{n+1}_{i,j}}{\sqrt{\varepsilon^2+(D_{0x}u^n_{i,j})^2+(D_{+y}u^n_{i,j})^2}} - \frac{u^{n+1}_{i,j}-u^n_{i,j-1}}{\sqrt{\varepsilon^2+(D_{0x}u^n_{i,j-1})^2+(D_{-y}u^n_{i,j})^2}}\right].
\end{aligned}$$

Introducing the notations

$$\begin{aligned}
c_E &:= \frac{1}{\sqrt{\varepsilon^2+(D_{+x}u^n_{i,j})^2+(D_{0y}u^n_{i,j})^2}}, \quad c_W := \frac{1}{\sqrt{\varepsilon^2+(D_{-x}u^n_{i,j})^2+(D_{0y}u^n_{i-1,j})^2}}, \\
c_S &:= \frac{1}{\sqrt{\varepsilon^2+(D_{0x}u^n_{i,j})^2+(D_{+y}u^n_{i,j})^2}}, \quad c_N := \frac{1}{\sqrt{\varepsilon^2+(D_{0x}u^n_{i,j-1})^2+(D_{-y}u^n_{i,j})^2}},
\end{aligned}$$

and solving for $u^{n+1}_{i,j}$, we obtain the iterative scheme

$$u^{n+1}_{i,j} = \frac{2\lambda h^2 f_{i,j} + c_E u^n_{i+1,j} + c_W u^n_{i-1,j} + c_S u^n_{i,j+1} + c_N u^n_{i,j-1}}{2\lambda h^2 + c_E + c_W + c_S + c_N}. \tag{4.7}$$

Using the most recent (north and west) values of $u_{i,j}$'s amounts to the Gauss–Seidel scheme which we use in the examples below for computation at all interior points $(x_i, y_j) \in \Omega$. The interior Gauss–Seidel scheme is augmented by reflection boundary conditions, in agreement with the Neumann boundary conditions (4.5). To

this end, we also reflect f outside Ω (by adding up to 10 gridlines on all sides of Ω). As the initial condition we set $u^0_{i,j} = f_{i,j}$.

In order to avoid grid effects, we rotate the starting point of the scheme (4.7) between the four corners of the grid, namely, $(0,0)$, $(i_{max},0)$, (i_{max},j_{max}), and $(0,j_{max})$, and alternate whether we run the algorithm row by row or column by column.

The scheme is iterated, $n = 0,1,\ldots,n_\infty$, until $\|u^{n_\infty} - u^{n_\infty-1}\|$ is reduced below a preassigned tolerance, so that $u^{n_\infty}_{i,j}$ produces an accurate approximation of the fixed point steady solution $u_\lambda(x_i,y_j)$. In general, $n_\infty = n_\infty(\lambda,h)$ is dictated by the contractivity of the fixed point iterations (4.7), e.g., in [6], while we note in passing that the following maximum principle holds: $\min_{i,j}|f_{i,j}| \le |u^n_{i,j}| \le \max_{i,j}|f_{i,j}|$ (in agreement with the maximum principle, $0 \le \min f \le u_\lambda(x) \le \max f$; see [19, section 1.14]).

This completes the description of the Euler–Lagrange scheme for a fixed λ. We tag the final discrete solution as $u_\lambda = \{u^{n_\infty}_{i,j}\}$. In order to convert this into the hierarchical multiscale decomposition, we reiterate this process, each time updating the value of f and λ in the following way:

$$\begin{cases} f_{new} \longleftarrow f_{current} - u_\lambda, \\ \lambda_{new} \longleftarrow 2\lambda_{current}. \end{cases}$$

In other words, we take the residual of the previous step and apply the $J(f_{current}-u_\lambda, 2\lambda)$ minimization using a doubled scaling parameter. With $\lambda_j = \lambda_0 2^j$, the final result after k steps is a multiscale representation of f, expressed in term of $u_j = u_{\lambda_j}$ and given by $f = u_0 + u_1 + u_2 + \cdots + u_k + v_k$.

How many hierarchical steps, k, should we take? Let us mention several stopping criteria. The first, measuring the amount of texture $\|v_{2^{k-1}}\|_{W^{-1,\infty}}$ to be below a certain tolerance factor, amounts to specifying the number of iterations, since in view of Theorem 2.1, $\|v_{2^k}\|_{W^{-1,\infty}} = 1/\lambda_{k+1}$. Another option would be to measure the energy, $\|u_k - u_{k-1}\|_{L^2(\Omega)}$, below a specified tolerance. The advantage of the hierarchical decomposition is that it also allows us to access the λ_k scale through the k-component, $\|u_k\|_{BV}/\lambda_k \sim \|u_k\|_{BV}/\lambda_k + \|u\|_2^2$; equivalently, the latter is a measure for the change in the L^2 texture, requiring $\|v_k\|_2^2 - \|v_{k-1}\|_2^2$ to be less than a specified tolerance factor.

We now turn to a series of numerical experiments which illustrate the hierarchical multiscale expansion for images. The different numerical results shown below use the same regularization parameter, $\varepsilon^2 = 10^{-6}$. We begin with a simple illustration for the improvement obtained by increasing the number of hierarchical iterations. In the simple case of a characteristic function of a disk (see Figure 1), the additional iterations improve the resolution as seen in the series of enhanced textures, $v_{\lambda_j} + 120$.

Next, we illustrate how the hierarchical decomposition of an image resolves detailed textures; see the increased scales of a fingerprint in Figure 2. In Figure 3 we illustrate hierarchical decomposition of a woman figure. In each hierarchical step, an additional amount of blurred texture is resolved in terms of the refined scaling for edges.

The following three figures zoom into one piece of the woman in Figure 3. In Figure 4, we see how our multiscale decomposition adds details of texture at each stage of the algorithm. Figures 5 and 6 show the different hierarchical stages—the u_j's and the v_j (enhanced by an additive factor of 120)—which add up to our final result in Figure 4. We record the first six terms since the remaining ones are not noticeable.

Next we turn to two numerical tests with noisy data. In Figure 7 we illustrate an additive noise. After 9 steps, the texture of the image is recovered on the top right

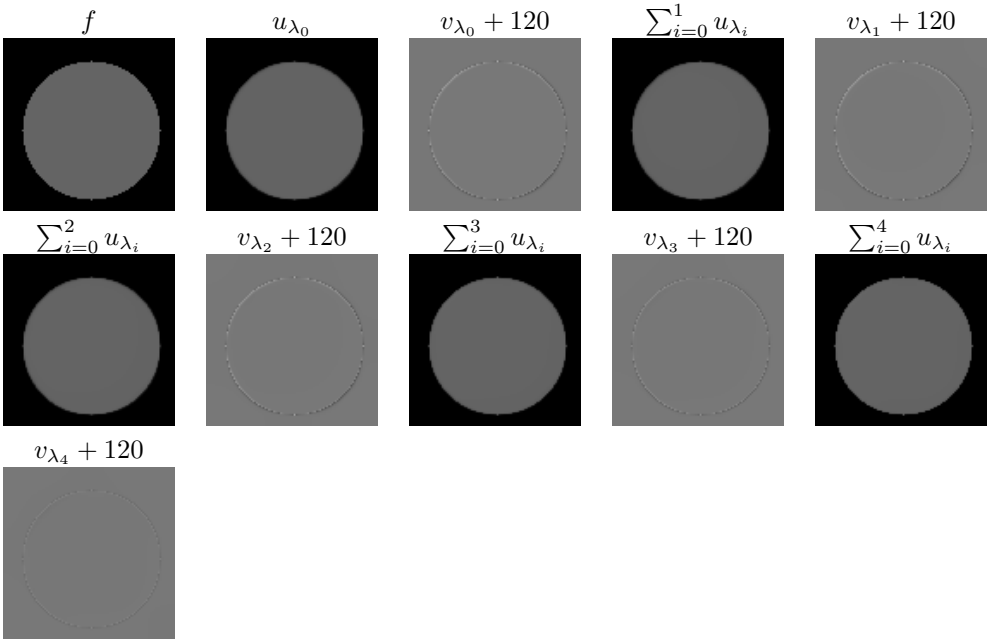

FIG. 1. *The u_{λ_j} components and the residuals, v_{λ_j}, for 5 steps, starting with an initial image of a circle,* (3.5). *Parameters: $\lambda_0 = .01$ and $\lambda_j = \lambda_0 2^j$.*

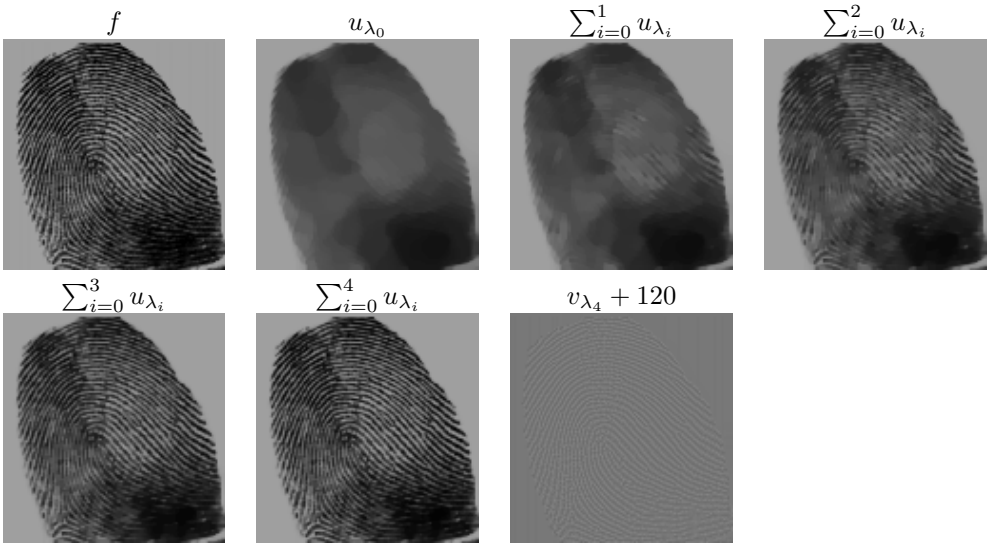

FIG. 2. *Decomposition of an initial image of a fingerprint for 5 steps with $\lambda_0 = .01$.*

corner of the image while removing a smaller scale noise from the woman forehead. If we continue the decomposition into smaller scales, then noise will reappear in the u components, as the refined scales reach the same scales of the noise itself. Figure 8 is another example of a noisy image. After 9 steps we obtain a denoised image while most of the texture is kept.

FIG. 3. *Successive decompositions of an image of a woman with* $\lambda_0 = .0005$.

The last examples demonstrated how the hierarchical decomposition separates between different features of edges, texture, and noise. The distinctive feature is their different scale. Our final example deals with different scales in an image of a galaxy shown in Figure 9. The smaller values of the scaling factor, λ, correspond to the larger objects in the image, while the smaller objects are brought into light when increasing values of λ are considered. In this manner, the hierarchical decomposition enables an effective *separation of scales* depicted, for example, in the last two images in Figure 9.

4.4. Localization of the hierarchical expansion. We want to localize the hierarchical algorithm so that most of the computational work concentrates in the neighborhoods of edges and textures while large homogeneous regions require a relatively smaller amount of work. To this end, we start by considering the whole domain embedded in a computational square $\Omega_0 := \Omega$. We then dyadically split each typical computational box, $\Omega_{j,k}$, into four new subregions, $\Omega_{j+1,k}$, depending on how much texture they have. We refer the reader to [11] for a similar adaptive approach where the local variation, $\|f - ave(f)\|_{L^2(\Omega_{j,k})}$, was used as a criterion for local refinement,

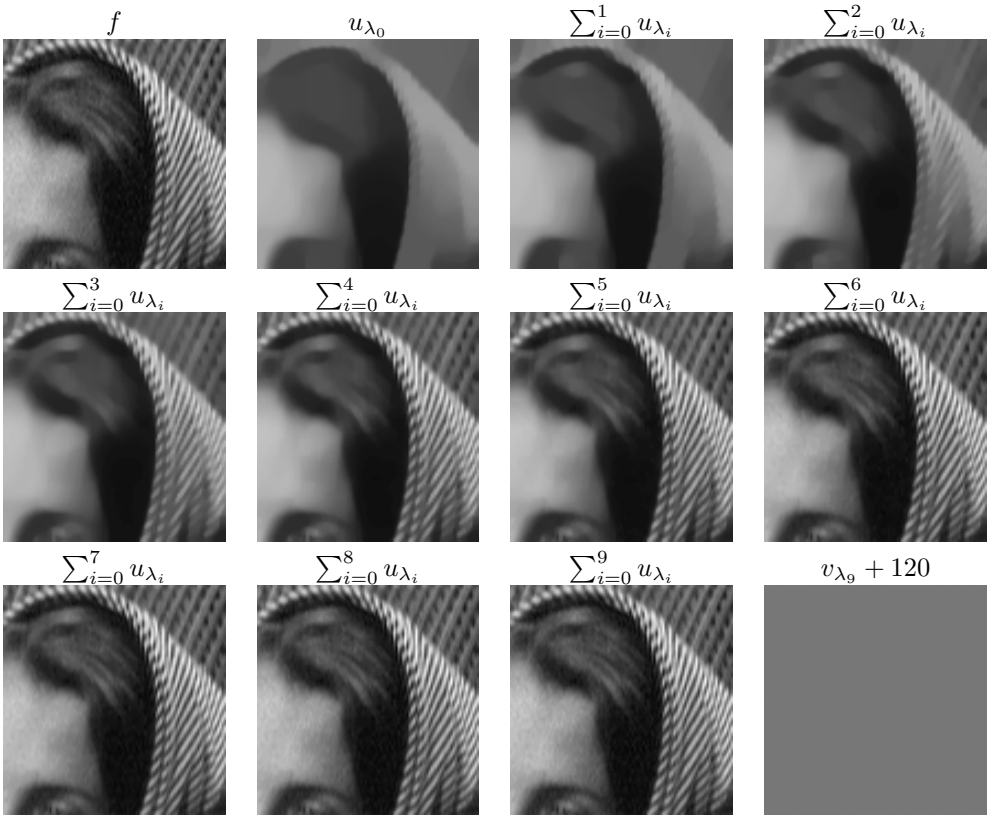

FIG. 4. *Decomposition of an initial image of a woman for* 10 *steps. Parameters:* $\lambda_0 = .005$ *and* $\lambda_j = \lambda_0 2^j$.

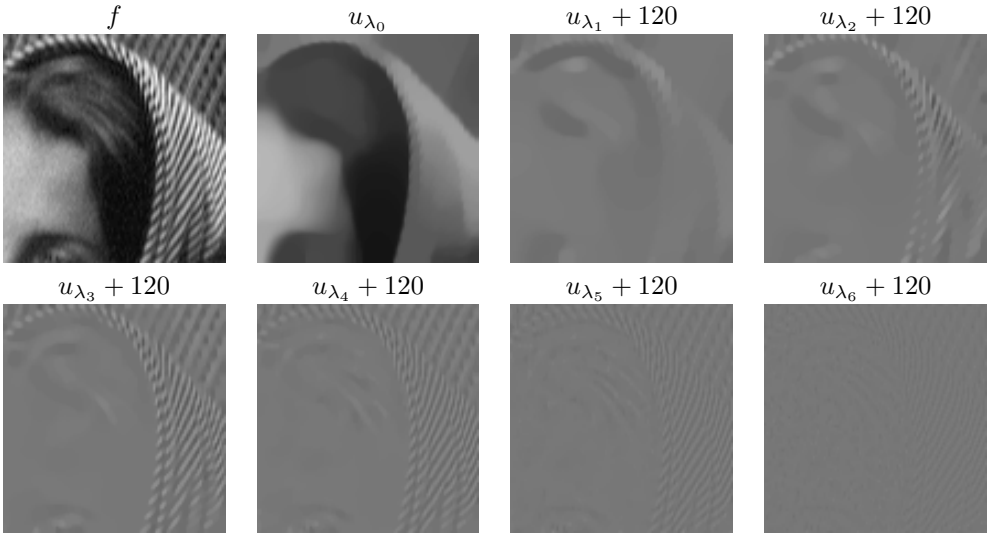

FIG. 5. *Representation of each* u_j *for* $0 \le j \le 6$. *Parameters:* $\lambda_0 = .005$ *and* $\lambda_j = \lambda_0 2^j$.

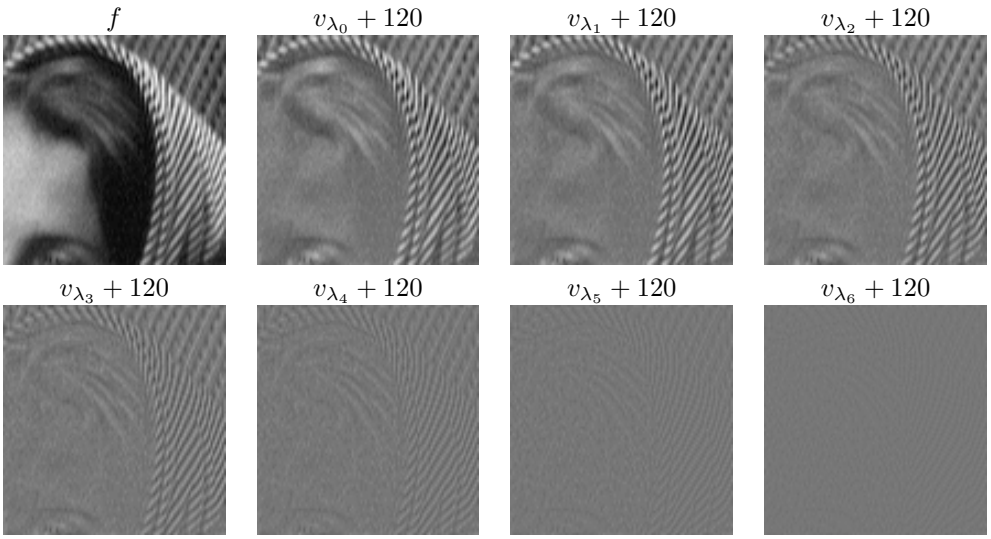

FIG. 6. *Representation of each* v_j *for* $0 \le j \le 6$. *Parameters:* $\lambda_0 = .005$ *and* $\lambda_j = \lambda_0 2^j$.

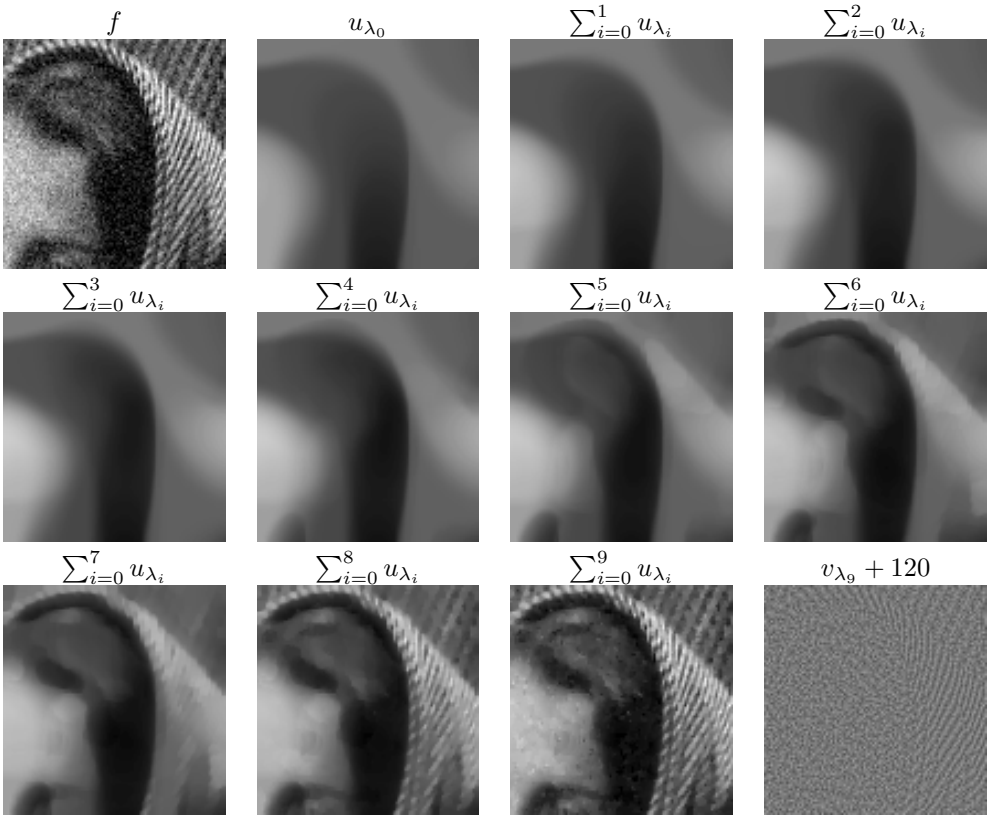

FIG. 7. *The recovery of* u *given an initial noisy image of a woman. Parameters:* $\lambda_0 = .0001$, $k = 10$, *and* $\lambda_k = \lambda_0 2^j$.

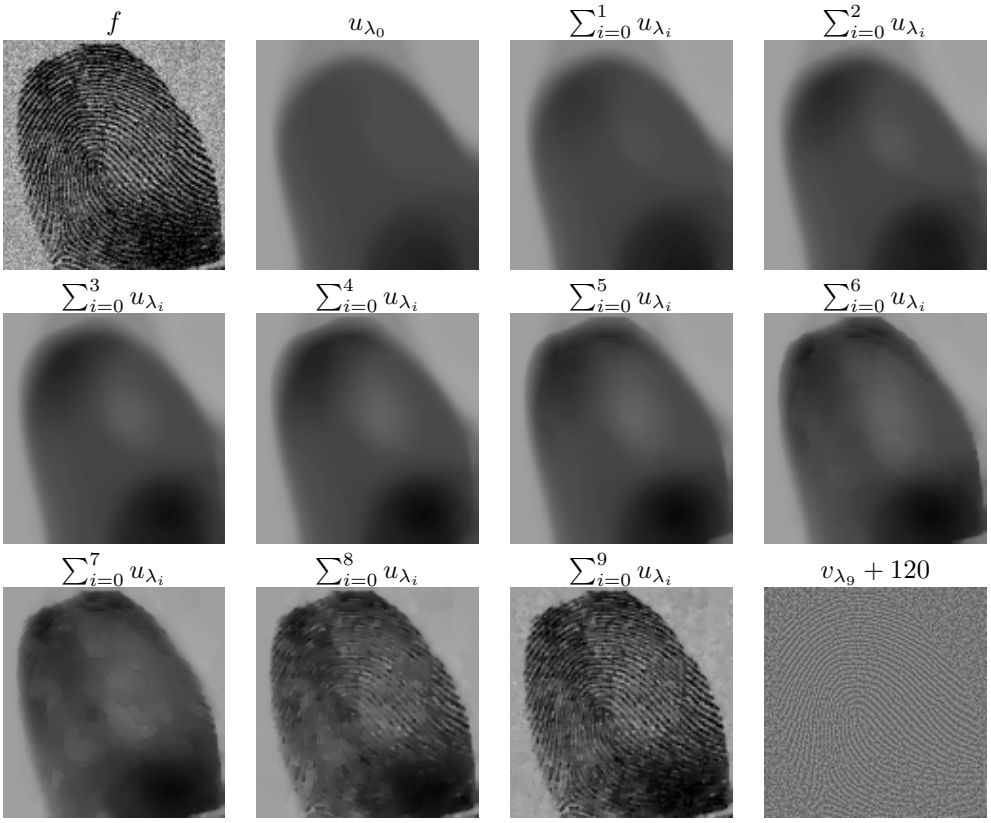

FIG. 8. *Decomposition of a noisy image of a fingerprint for* 10 *steps. Parameters:* $\lambda_0 = .0001$ *and* $\lambda_j = \lambda_0 2^j$.

based on equidistribution of local variations. In the present context of hierarchical decompositions, we propose two different refinement criteria to decide whether to stop the refinement of the current box $\Omega_{j,k}$:

(i) The BV norm of the local residual—the refinement continues until $\|v_\lambda\|_{BV(\Omega_{j,k})}$ is below a given tolerance factor, δ.

(ii) A weaker stopping criterion based on the value of the localized minimizer—if $J(v_j, \lambda_j)_{\Omega_{j,k}} \le \delta$, then the refinement stops. Comparing the optimal pair $[u_{j+1}, v_{j+1}]$ with the trivial decomposition $[v_j, 0]$ implies that the first criterion is indeed stronger for $J(v_j, \lambda_j) \le \|v_j\|_{BV}$. In practice, however, the numerical results presented below show the two refinement criteria yield similar results.

Let us describe the details for the second adaptive procedure with a typical example of an image of $2^m \times 2^m$ pixels (the initial size of an image is always extended to next dyadic size by reflection). We let Ω_0 denote this initial computational domain, and we recall that at each stage the computational boxes need to be padded with five additional rows on each side to implement reflection boundary condition. If $J(f, \lambda, BV(\Omega_0), L^2(\Omega_0)) \le \delta$, then we pursue the hierarchical decomposition of f in Ω_0. Otherwise, if $J(f, \lambda, BV(\Omega), L^2(\Omega)) \ge \delta$, we split the initial $2^m \times 2^m$ region into four equal images, each of $2^{m-1} \times 2^{m-1}$ pixels; see Figure 10. They are augmented with an extended boundary of five rows on each side, making four computational boxes, $\Omega_{1,k}$, $k = 1, \dots 4$, each of $(2^{m-1} + 10) \times (2^{m-1} + 10)$ pixels. Note that there is

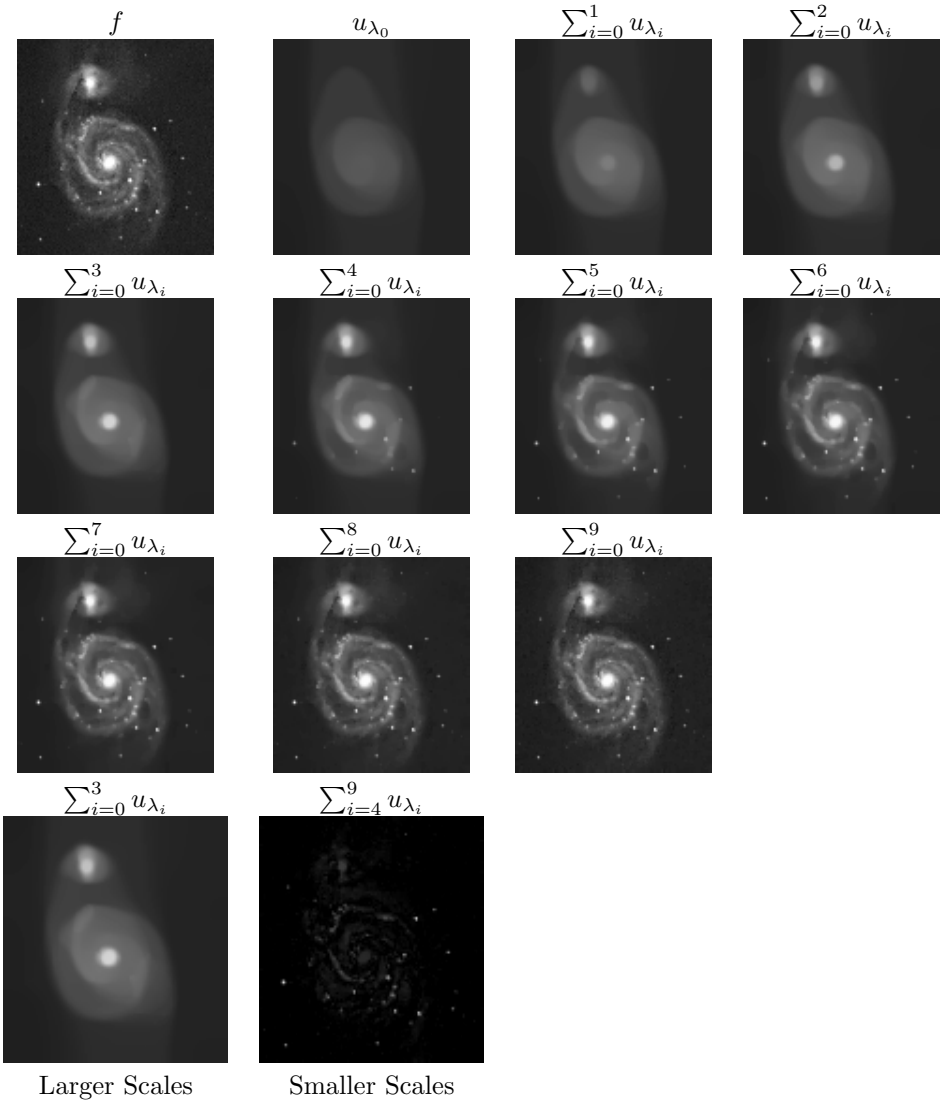

FIG. 9. *Decomposition of an image of a galaxy for* 10 *steps. Parameters:* $\lambda_0 = .001$ *and* $\lambda_j = \lambda_0 2^j$. *The last two images illustrate separation of scales.*

overlapping between the new four images due to the reflection boundary conditions. These artificial boundaries are ignored when subregions are pieced together to avoid obvious lines along different zones where splitting took place. Now, for each of the new four subregions, $\Omega_{1,k}$, we first check whether $J(f, \lambda; BV(\Omega_{1,k}), L^2(\Omega_{1,k})) \leq \delta$, and we continue the refinement until either the value of J becomes smaller than the tolerance δ or we reach the smallest boxes of 2×2 pixels. In each computational box $\Omega_{j,k}$ satisfying the refinement stopping criterion, we pursue our multiscale decomposition seeking the minimizing pairs $[u_{j+1}, v_{j+1}]$ of $J(v_j, \lambda_{j+1}, BV(\Omega_{j,k}), L^2(\Omega_{j,k}))$. While iterating these hierarchical stages, we check whether $J(v_j, \lambda_{j+1}) \leq \delta$: if we do not satisfy the desired tolerance at this point, we continue with the splitting process; if we do, then we continue with our hierarchical expansion.

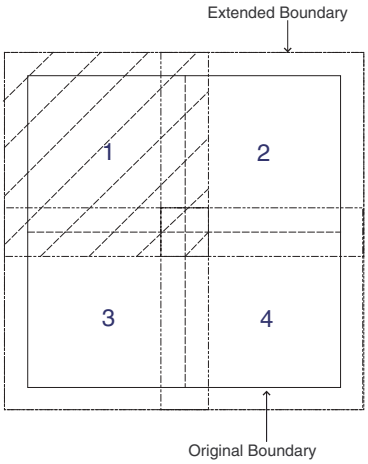

FIG. 10. *Method of splitting the domain into four equal regions. The shaded region represents the first image.*

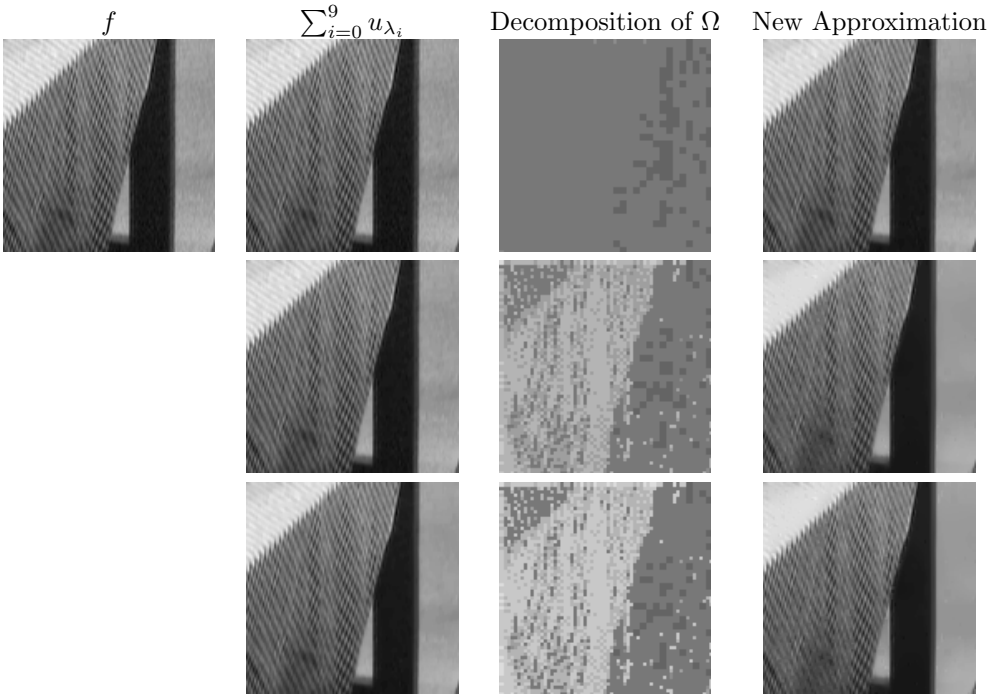

FIG. 11. *Adaptive decomposition of f and Ω using $J(u_\lambda)$ as a refinement criterion. We use $\delta = 50 \times 128^2$ as a tolerance threshold for all calculations. Parameters: Row* 1: $\lambda_0 = .01$, *Row* 2: $\lambda_0 = .001$, *and Row* 3: $\lambda_0 = .0005$, *where* $\lambda_j = \lambda_0 2^j$.

In Figure 11, we consider $J(v_j)$, and in Figure 12, we use $\|v_j\|_{BV}$. The first column represents the general, nonadaptive hierarchical algorithm. The middle column represents the adaptive refinements of Ω where lighter indicates more texture. The last column contains the new adaptive expansion. The resulting image u is therefore

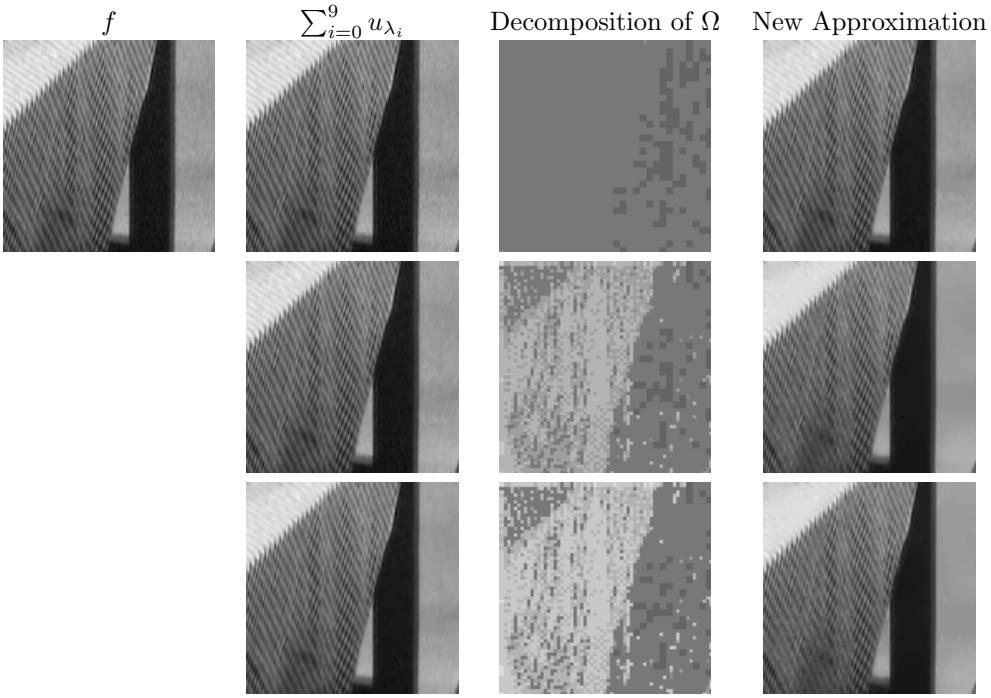

FIG. 12. *Adaptive decomposition of f and Ω using the total variation of v_λ as a refinement criterion. We use $\delta = 50 \times 128^2$ as a tolerance threshold for all calculations. Parameters: Row* 1: $\lambda_0 = .01$, *Row* 2: $\lambda_0 = .001$, *and Row* 3: $\lambda_0 = .0005$, *where* $\lambda_j = \lambda_0 2^j$.

comprised of some regions which required as much as 10 terms, while others need only 3 terms. What is remarkable is how close the adaptive approximation is to the full algorithm, even with a fairly large tolerance δ. Also, both stopping criteria, based on $\|v_j\|_{BV} \le \delta$ and $J(v_j, \lambda_{j+1}) \le \delta$, yield similar results.

4.5. The hierarchical (BV, L^2) decomposition scheme for color images. We record here the formulae for color images, which are realized in terms of vector-valued functions $\mathbf{f} = (f_1, f_2, f_3) \in L^2(\mathbb{R}^2)^3$. The corresponding minimizer for color image restoration of Rudin, Osher, and Fatemi [25] reads

$$J(\mathbf{f}, \lambda) = \inf_{\mathbf{u} \in BV} \Big\{ \lambda \|\mathbf{f} - \mathbf{u}\|_{L^2}^2 + \|\mathbf{u}\|_{BV} \Big\}.$$

Here, the BV and L^2 norms of the corresponding 3-vectors, $\mathbf{u} = (u_1, u_2, u_3)$ and $\mathbf{v} = \mathbf{f} - \mathbf{u} = (v_1, v_2, v_3)$, are defined in terms of their Euclidean structure:

$$\|\mathbf{u}\|_{BV} := \sup_{\varphi \in C_0^\infty} \left\{ \int \langle \mathbf{u}, \nabla \varphi \rangle \mid \|\varphi\|_{L^\infty} \le 1 \right\}, \quad \|\mathbf{v}\|_{L^2}^2 = \int |\mathbf{v}|^2 dx.$$

Formally minimizing the above energy with respect to u_1, u_2, and u_3 yields the following Euler–Lagrange system of coupled PDEs:

(4.8)
$$u_\ell - \frac{1}{2\lambda} \mathrm{div}\left(\frac{\nabla u_\ell}{|\nabla \mathbf{u}|} \right) = f_\ell, \quad \ell = 1, 2, 3, \qquad |\nabla \mathbf{u}| = \sqrt{|\nabla u_1|^2 + |\nabla u_2|^2 + |\nabla u_3|^2}.$$

Its approximate solution is computed through an iterative scheme similar to the previous scalar case, except the need to solve three coupled equations for the 3-vector of unknowns, $\mathbf{u}^n_{i,j} = ((\mathbf{u}_\ell)^n_{i,j})$, $\ell = 1, 2, 3$, with the corresponding regularized gradients

$$\begin{aligned}
c_{\ell;E} &= \frac{1}{\sqrt{\varepsilon^2 + (D_{+x}(\mathbf{u}_\ell)^n_{i,j})^2 + (D_{0y}(\mathbf{u}_\ell)^n_{i,j})^2}},\\
c_{\ell;W} &:= \frac{1}{\sqrt{\varepsilon^2 + (D_{-x}(\mathbf{u}_\ell)^n_{i,j})^2 + (D_{0y}(\mathbf{u}_\ell)^n_{i-1,j})^2}},\\
c_{\ell;S} &:= \frac{1}{\sqrt{\varepsilon^2 + (D_{0x}(\mathbf{u}_\ell)^n_{i,j})^2 + (D_{+y}(\mathbf{u}_\ell)^n_{i,j})^2}},\\
c_{\ell;N} &:= \frac{1}{\sqrt{\varepsilon^2 + (D_{0x}(\mathbf{u}_\ell)^n_{i,j-1})^2 + (D_{-y}(\mathbf{u}_\ell)^n_{i,j})^2}}.
\end{aligned}$$

Solving for $(\mathbf{u}_\ell)^{n+1}_{i,j}$, we obtain the vector-valued iteration scheme

$$(\mathbf{u}_\ell)^{n+1}_{i,j} = \frac{2\lambda h^2(\mathbf{f}_\ell)_{i,j} + c_{\ell;E}(\mathbf{u}_\ell)^n_{i+1,j} + c_{\ell;W}(\mathbf{u}_\ell)^n_{i-1,j} + c_{\ell;S}(\mathbf{u}_\ell)^n_{i,j+1} + c_{\ell;N}(\mathbf{u}_\ell)^n_{i,j-1}}{2\lambda h^2 + c_{\ell;E} + c_{\ell;W} + c_{\ell;S} + c_{\ell;N}}, \quad \ell = 1, 2, 3. \tag{4.9}$$

Figure 13 demonstrates the hierarchical decomposition for a vector-valued MRI image in RGB mode.

5. Extensions. The hierarchical (BV, L^2) decomposition was introduced in the context of the J-minimizer of Rudin, Osher, and Fatemi [25]. It applies to other variational models arising in image analysis, and we shall mention three examples: a BV minimizer weighted by the presence of a *blurring* operator, a *multiplicative* version of the J-minimization (which is adapted for multiplicative rather than additive denoising), and the Mumford–Shah segmentation model [21] which for the purpose of our computations is realized by the elliptic regularization of Ambrosio and Tortorelli [3]. We briefly discuss the hierarchical decompositions in these three models below. Preliminary results are found in [22], and a more detailed account will be provided in a forthcoming paper.

5.1. Hierarchical decomposition of blurred images. Given $f \in L^2(\Omega)$, a cut-off parameter $\lambda > 0$, and a blurring kernel K (a linear and continuous operator from $L^2(\Omega)$ to $L^2(\Omega)$), we consider a decomposition of f provided by the following $J_K(f, \lambda)$ minimization in the presence of blur; see, e.g., [19, section 1.14], [15]:

$$J_K(f, \lambda; BV, L^2) := \inf_{u \in BV} \Big\{ \lambda \|f - Ku\|^2_{L^2(\Omega)} + \|u\|_{BV(\Omega)} \Big\}. \tag{5.1}$$

Let $v_\lambda := f - Ku_\lambda$ denote the "texture" so that $f = Ku_\lambda + v_\lambda$. Starting with $\lambda = \lambda_0$ in (5.1) we proceed, as before, iterating the hierarchical decomposition, $v_j = Ku_{j+1} + v_{j+1}$, at scale $\lambda_j = \lambda_0 2^j$. We end up with hierarchical expansion of the blurred image f,

$$f = Ku_0 + Ku_1 + \cdots + Ku_{k-1} + Ku_k + v_k, \quad \lambda_j := \lambda_0 2^j,$$

which in turn paves the way for hierarchical, multiscale denoised representation $\sum_j^k u_j$. In particular, the total energy of (sufficiently smooth) f is decomposed into its dyadic

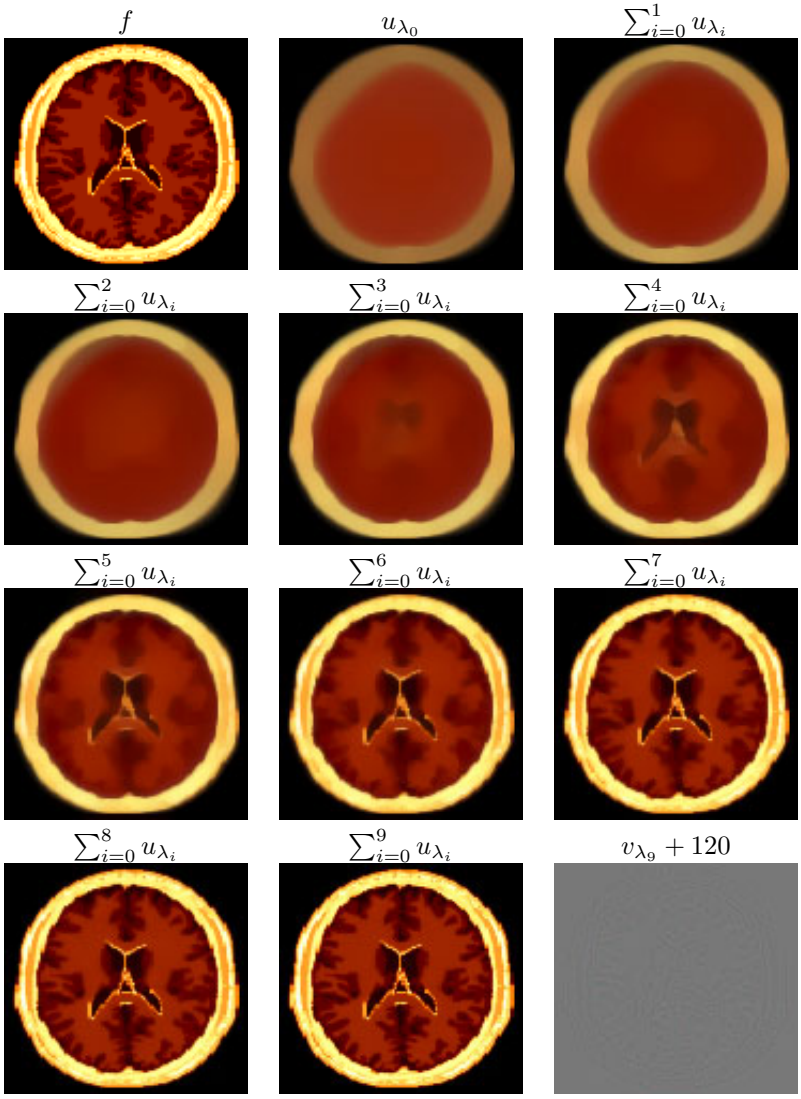

FIG. 13. *Decomposition of a vector-valued MRI image for* 10 *steps. Parameters:* $\lambda_0 = .00025$ *and* $\lambda_k = \lambda_0 2^j$.

building blocks,

$$\|f\|_2^2 = \sum_{j=0}^{\infty} \left[\frac{1}{\lambda_j} \|u_j\|_{BV} + \|Ku_j\|_2^2 \right].$$

Arguing along the lines of (2.16) we find that the last statement has an equivalent energy decomposition, $\|f\|_2^2 \sim \sum_j \|u_j\|_{BV}/\lambda_j$, where the *explicit* dependence on the blur K is removed.

5.2. Hierarchical decomposition of images with multiplicative noise. Following [26], we consider a *multiplicative* degradation model where we are given a blurred image $f = u \cdot v$, with $u > 0$ being the original image and where v models

the multiplicative noise, normalized such that $\int_\Omega v(x,y)dxdy = 1$. Let u_λ be the minimizer of the corresponding total variation functional in the multiplicative case [26]

$$M(f,\lambda;BV,L^2) := \inf_{u\in BV_+(\Omega)} \left\{ \lambda \left\| \frac{f}{u} - 1 \right\|^2_{L^2(\Omega)} + \|u\|_{BV(\Omega)} \right\}. \tag{5.2}$$

Setting $v_\lambda := \frac{f}{u_\lambda}$ we end up with the one-scale decomposition $f = u_\lambda v_\lambda$. We construct the hierarchical decomposition as before, except that sums and differences are replaced by products and divisions. Thus the iterative step at scale λ_j reads $v_j = v_{j+1}u_{j+1}$, leading to the multiplicative hierarchical decomposition

$$f = u_0, u_1, \dots, u_k \times v_k, \qquad \lambda_j = \lambda_0 2^j.$$

5.3. The hierarchical (SBV, L^2) decomposition. We want to apply the hierarchical decomposition to the Mumford and Shah functional [21]. To this end we consider its elliptic approximation of Ambrosio and Tortorelli [3],

$$AT^\varepsilon(f,\lambda) := \inf_{\{w,u,v \,|\, u+v=f\}} \left\{ \int_\Omega \Big[w^2|\nabla u|^2 + |v|^2 \Big] dx + \lambda \left[\varepsilon \|\nabla w\|^2_{L^2} + \frac{\|w-1\|^2_{L^2}}{\varepsilon} \right] \right\}.$$

Let $[u_\lambda, v_\lambda]$ be the minimizer of $AT^\varepsilon(f,\lambda)$ (depending on w). Here f is modeled as a u_λ which is restricted to the smaller SBV space (a special subclass of BV space, consisting of measure gradients free of the Cantor component [3]), while the texture v_λ lives in L^2. We proceed to construct the hierarchical (SBV, L^2) decomposition of f in the same manner as before, letting $[u_{j+1}, v_{j+1}]$ be the AT minimizer

$$[u_{j+1}, v_{j+1}] = \operatorname*{arginf}_{u+v=f} AT^\varepsilon(v_j, \lambda_j), \quad \lambda_j = \lambda_0 2^j.$$

We end up with the hierarchical decomposition

$$f = u_0 + u_1 + \cdots + u_k + v_k.$$

Here, at each hierarchical step, we also obtain the edge detectors $1 - w_j = 1 - w_{\lambda_j}$, which are (essentially) supported along the boundaries of objects enclosed by edges identified by u_j.

REFERENCES

[1] R. Acar and C. R. Vogel, *Analysis of bounded variation penalty methods of ill-posed problems*, Inverse Problems, 10 (1994), pp. 1217–1229.

[2] L. Ambrosio, N. Fusco, and D. Pallara, *Functions of Bounded Variation and Free Discontinuity Problems*, Oxford University Press, New York, 2000.

[3] L. Ambrosio and V. Tortorelli, *On the approximation of functionals depending on jumps by elliptic functionals via* Γ*-convergence*, Comm. Pure Appl. Math., 43 (1990), pp. 999–1036.

[4] F. Andreu, C. Ballester, V. Casselles, and J. M. Mazon, *Minimizing total variation flow*, C. R. Acad. Sci. Paris Sér. I Math., 331 (2000), pp. 867–872.

[5] G. Aubert and P. Kornprobst, *Mathematical Problems in Image Processing*, Springer-Verlag, New York, 2002.

[6] G. Aubert and L. Vese, *A variational method in image recovery*, SIAM J. Numer. Anal., 34 (1997), pp. 1948–1979.

[7] A. Z. Averbuch, R. R. Coifman, F. G. Meyer, and J. O. Strömberg, *Multi-layered image representation: Application to image compression*, in Proceedings of the International Conference on Image Processing (ICIP 1998), Vol. 2, Chicago, IL, 1998, IEEE Computer Society Press, Los Alamitos, CA, 1998, pp. 292–296.

[8] J. Bergh and J. Lofstrom, *Interpolation Spaces. An Introduction*, Springer-Verlag, New York, 1976.
[9] C. Bennett and R. Sharpley, *Interpolation of Operators*, Academic Press, Boston, MA, 1988.
[10] A. Chambolle and P. L. Lions, *Image recovery via total variation minimization and related problems*, Numer. Math., 76 (1997), pp. 167–188.
[11] A. Cohen, R. DeVore, P. Petrushev, and H. Xu, *Nonlinear approximation and the space $BV(\mathbb{R}^2)$*, Amer. J. Math., 121 (1999), pp. 587–628.
[12] A. Cohen, W. Dahmen, I. Daubechies, and R. DeVore, *Harmonic analysis of the space BV*, Rev. Mat. Iberoamericana, 19 (2003), pp. 235–263.
[13] R. DeVore and G. Lorentz, *Constructive Approximation*, Grundlehren Math. Wiss. 303, Springer-Verlag, Berlin, 1993.
[14] R. DeVore and B. Lucier, *Wavelets*, in Acta Numerica 1, Cambridge University Press, Cambridge, UK, 1992, pp. 1–56.
[15] D. Donoho, *Nonlinear solution of linear inverse problems by wavelet-vaguelette decomposition*, Appl. Comput. Harmon. Anal., 2 (1995), pp. 101–126.
[16] L. C. Evans and R. F. Gariepy, *Measure Theory and Fine Properties of Functions*, CRC Press, London, 1992.
[17] Y. Gousseau and J.-M. Morel, *Are natural images of bounded variation?*, SIAM J. Math. Anal., 33 (2001), pp. 634–648.
[18] F. G. Meyer, A. Z. Averbuch, and R. R. Coifman, *Multilayered image representation: Application to image compression*, IEEE Trans. Image Process., 11 (2002), pp. 1072–1080.
[19] Y. Meyer, *Oscillating Patterns in Image Processing and Nonlinear Evolution Equations*, Univ. Lecture Ser. 22, AMS, Providence, RI, 2002.
[20] M. Mumford and B. Gidas, *Stochastic models for generic images*, Quart. Appl. Math., 51 (2001), pp. 85–111.
[21] M. Mumford and J. Shah, *Boundary detection by minimization functionals*, in Proceedings of the IEEE Computer Vision Pattern Recognition Conference, San Francisco, CA, 1985.
[22] S. Nezzar, *A Multiscale Image Representation Using Hierarchical (BV, L^2) Decompositions*, Ph.D. Thesis, University of California, Los Angeles, CA, 2003.
[23] L. Rudin, *Images, Numerical Analysis of Singularities, and Shock Filters*, Ph.D. thesis, Computer Science Department, California Institute of Technology, Pasadena, CA, 1987.
[24] L. Rudin and V. Caselles, *Image recovery via multiscale total variation*, in Proceedings of the Second European Conference on Image Processing, Palma, Spain, 1995.
[25] L. Rudin, S. Osher, and E. Fatemi, *Nonlinear total variation based noise removal algorithms*, Phys. D, 60 (1992), pp. 259–268.
[26] L. Rudin and S. Osher, *Total variation based image restoration with free local constraints*, in Proceedings of the International Conference on Image Processing (ICIP 1994), Vol. I, Austin, TX, 1994, IEEE Computer Society Press, Los Alamitos, CA, 1994, pp. 31–35.
[27] L. Vese, *A study in the BV space of a denoising-deblurring variational problem*, Appl. Math. Optim., 44 (2001), pp. 131–161.