SIAM J. IMAGING SCIENCES
Vol. 4, No. 1, pp. 300–312
© 2011 Society for Industrial and Applied Mathematics

Integro-Differential Equations Based on (BV, L^1) Image Decomposition*

Prashant Athavale† and Eitan Tadmor‡

Abstract. A novel approach for multiscale image processing based on integro-differential equations (IDEs) was proposed in [E. Tadmor and P. Athavale, *Inverse Probl. Imaging*, 3 (2009), pp. 693–710]. These IDEs, which stem naturally from multiscale (BV, L^2) hierarchical decompositions, yield inverse scale representations of images in the sense that the BV-dual norms of their residuals are inversely proportional to the scaling parameters. Motivated by the fact that (BV, L^1) decomposition is more suitable for extracting *local* scale-space features than (BV, L^2), we introduce here the IDEs which arise from multiscale (BV, L^1) hierarchical decompositions. We study several variants of this (BV, L^1)-based IDE model, depending on modifications to the curvature term.

Key words. hierarchical image decompositions, total variation, denoising, deblurring, inverse scale, variational problem, integro-differential equation, energy decomposition

AMS subject classifications. 26B30, 65C20, 68U10, 92C50

DOI. 10.1137/100795504

1. Introduction.

In [32] we introduced a class of integro-differential equations (IDEs) to model images, based on (BV, L^2) hierarchical decompositions [33, 34]. In this paper we study integro-differential equations arising from hierarchical (BV, L^1) image decomposition. As in [32], we work here with grayscale images which can be realized as a graph of a discrete function $f : \Omega \subset \mathbb{R}^2 \to \mathbb{R}$. The values of this function, $f(x)$, denote the intensity of the image at the discrete points $x \in \Omega$: the function f attains its maximum value at the brightest spots in the image and its minimum value of zero at the darkest spots. The graph of an image consists of discrete pixels, which, for mathematical analysis, is postulated as an $L^2(\Omega)$ function. The image f, which is obtained by any image capturing device, may be blurred and may contain some additive noise η. Thus, the observed image, f, could be written as $f = KU + \eta$, where U is the clean image sought without blurring and noise and $K : L^2(\Omega) \to L^2(\Omega)$ is some blurring operator. The problem of obtaining the "clean" image U from the observed image f is an ill-posed problem of image restoration, which can be addressed by several inverse problem solvers. We mention in this context variational techniques using Tikhonov-like regularization, PDE-based methods, filtering, stochastic modeling, and wavelet-based techniques that were developed for solving these image processing problems [5, 9, 14, 15, 17, 18, 20, 21, 24, 25, 29, 30, 35].

*Received by the editors May 17, 2010; accepted for publication (in revised form) January 13, 2011; published electronically March 17, 2011. This research was supported in part by NSF grants DMS07-07949 and DMS10-08397.
http://www.siam.org/journals/siims/4-1/79550.html

†Institute of Pure and Applied Mathematics, University of California, Los Angeles, CA 90095 (prashant@math.ucla.edu).

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

Image restoration leads to *image decomposition*. For example, any denoising of an observed image f results in the decomposing of type $f = U_\lambda + \eta_\lambda$, where η_λ is interpreted as a noise by the denoising method. Here, λ is an algorithm-specific *scaling parameter*: in the case of Gaussian smoothing, for example, the variance of the Gaussian kernel may serve as such a scaling parameter. Small-scale features, categorized as noise, are then forced into η_λ, resulting in a larger-scale version, U_λ, of the original image f.

1.1. PDE-based, variational, and hierarchical decompositions. Two main approaches for multiscale representations of images are PDE-based methods and variational methods. The basic PDE-based method is the heat equation, $u_t = \Delta u$, where, starting with the given image as initial condition, $u(0) := f$, it produces a multiscale representation, $\{u(t)\}_{t>0}$, for the image f. The heat equation removes noise through blurring, but being an isotropic diffusion it also blurs edges which define main features in the image. This drawback can be removed by using instead the Perona–Malik [29] nonlinear equation based on *nonisotropic* diffusion, $u_t = \operatorname{div}(g(|\nabla u|)\nabla u)$. The equation is ill-posed, however, and as already noted by these authors, the Perona–Malik diffusion can lead to false detection of edges in the presence of noise. To resolve this issue, Catté et al. [11] proposed using a modified nonisotropic and nonlocal diffusive amplitude of the form $g(|G_\sigma \star \nabla u|)$; this class of methods was studied in [19, 36]. All these methods give rise to a *forward* multiscale representation of the image f, in the sense that $u(t)$ begins with the finest scale, $u(0) := f$, and then progressively blurs into coarser representations of f as $t \to \infty$.

The class of variational methods is a widely used alternative to PDE-based methods in image processing. Methods such as Mumford–Shah segmentation [25, 26] and Rudin–Osher–Fatemi (ROF) decomposition [30] fall under a general category of Tikhonov regularization [35, 31]. Here one attempts to find a close approximation to a function $f \in Y$, in a space $X \subsetneq Y$, which is an appropriate space adapted to measure edges and textures sought in u. This leads to the following minimization problem:

$$f = u_\lambda + v_\lambda, \quad [u_\lambda, v_\lambda] := \operatorname*{arginf}_{f=u+v} \big\{ \|u\|_X + \lambda \|v\|_Y \big\}.$$

The term $\|u\|_X$ is a regularizing term, and $u_\lambda + v_\lambda$ is a multiscale decomposition of f which varies with the positive scaling parameter, λ. For references on this class of "$u + v$" methods we refer the reader to [10]. In the case of the ROF model [30], for example, edges are sought in the space of bounded variations, $X = BV(\Omega)$ and $f \in Y = L^2$; see, e.g., [6]. This yields the (BV, L^2)-decomposition:

$$f = u_\lambda + v_\lambda, \quad [u_\lambda, v_\lambda] := \operatorname*{arginf}_{f=u+v} \big\{ |u|_{BV} + \lambda \|v\|_{L^2}^2 \big\}, \tag{1.1}$$

where $|u|_{BV} := \int_\Omega |\nabla u|$ is the BV-seminorm. For small values of λ, the minimizer u_λ is a large-scale image which consists of only main features and prominent edges in f. As λ becomes larger, u_λ captures the small-scale details of f. The point of view advocated in [33] is to replace the fixed scaling parameter λ by a varying sequence of *inverse-scale* parameters, $\lambda_1 < \lambda_2 < \cdots$, which dictate the scaling of the iterative refinement step,

$$[u_{\lambda_{j+1}}, v_{\lambda_{j+1}}] := \operatorname*{arginf}_{v_{\lambda_j}=u+v} \big\{ |u|_{BV} + \lambda_{j+1} \|v\|_{L^2}^2 \big\}, \tag{1.2a}$$

which in turn generates a multiscale (BV, L^2) hierarchical representation,

$$f \approx \sum_j u_{\lambda_j}. \tag{1.2b}$$

2. A novel (BV, L^1)-based integro-differential equation (IDE). In [32] we introduced the integro-differential equation[1]

$$\int_0^t u(x,s)\,ds = f(x) + \frac{1}{2\lambda(t)} \operatorname{div}\left(\frac{\nabla u(x,t)}{|\nabla u(x,t)|}\right), \qquad \frac{\partial u}{\partial \mathbf{n}}\Big|_{\partial\Omega} = 0. \tag{2.1}$$

This novel IDE was based on a hierarchical (BV, L^2) decomposition (1.2) which was introduced by Tadmor, Nezzar, and Vese [33, 34], from which it inherits many important properties. Once we were dealing with the framework of IDEs, several extensions of (2.1) were proposed in [32] to deal with denoising and deblurring of images, which were not strictly associated with any variational problem.

The ROF decomposition (1.1) uses the L^2-fidelity term $\|f-u\|_{L^2}^2$. An alternative variational problem based on an L^1-fidelity term, $\|f-u\|_{L^1}$, was proposed earlier by Alliney [4] and was studied extensively by Chan, Esedoḡlu, and Nikolova (see [12, 27, 13]):

$$f = u_\lambda + v_\lambda, \quad [u_\lambda, v_\lambda] := \operatorname*{arginf}_{f=u+v} \left\{ |u|_{BV} + \lambda \|v\|_{L^1} \right\}. \tag{2.2}$$

The resulting (BV, L^1) minimization[1] differs from the (BV, L^2) model in several important aspects which have attracted considerable attention in recent years; see [1, 2, 3, 16]. We shall mention two such aspects. The (BV, L^1) minimization is *contrast invariant*, as opposed to the (BV, L^2) minimization. Chan and Esedoḡlu [12] also showed the more *local* geometric aspects of the (BV, L^1) minimization. In particular, recall that if $\operatorname{supp}(f) \in B_R(0)$, then both the (BV, L^1) and the (BV, L^2) minimizers vanish, namely, $u_\lambda \equiv 0$ and $v_\lambda = f$ for small λ's below the minimal threshold $\lambda < \lambda_L \propto \frac{1}{R}$. On the other hand, if the f is a characteristic function χ_Σ, then the (BV, L^1) minimizer of (2.2) admits a maximal threshold, λ_H, such that for all $\lambda > \lambda_H$ we have $u_\lambda = f$ and $v_\lambda \equiv 0$, in contrast to the "leaking" phenomenon with the (BV, L^2)-based ROF model, where $v_\lambda \propto \frac{1}{R}$ [22].

2.1. Quantization. To proceed with the derivation of a (BV, L^1)-based IDE, we first need to address a proper quantization of the image intensity. If we let τ denote the small intensity quanta, then we can rescale the coarse image in (2.2), u_λ, in τ-units. With this in mind, we decompose the given image f using the (BV, L^1) scheme, starting with initial scale λ_0:

$$f = u_{\lambda_0}\tau + v_{\lambda_0}, \quad u_{\lambda_0} := \operatorname*{arginf}_u \left\{ |u|_{BV} + \frac{\lambda_0}{\tau}\|f - u\tau\|_{L^1} \right\}.$$

[1]The term $\operatorname{div}\left(\frac{\nabla u(\cdot,t)}{|\nabla u(\cdot,t)|}\right)$, which often arises in image processing, is the curvature of isolevel curves of function $u(\cdot,t)$ [8, 7]. In the present context, it arises as the first variation of the BV-norm, $|u|_{BV}$, which in practice is replaced by its regularized form, e.g., $|u|_{BV} = \sqrt{|\nabla u|_{L^1}^2 + \varepsilon^2}$, thus removing the ambiguity when $\nabla u = 0$ and the possible lack of uniqueness in the corresponding (BV, L^1) minimization in (2.2) [1, 3].

The residual image $v_{\lambda_0} := f - u_{\lambda_0}\tau$ can be further decomposed into a smaller scale with $\lambda_1 > \lambda_0$:

$$v_{\lambda_0} = u_{\lambda_1}\tau + v_{\lambda_1}, \quad u_{\lambda_1} := \operatorname*{arginf}_u \left\{ |u|_{BV} + \frac{\lambda_1}{\tau}\|v_{\lambda_0} - u\tau\|_{L^1} \right\}.$$

We can continue this process for $\lambda_0 < \lambda_1 < \lambda_2 < \cdots$:

$$v_{\lambda_{k-1}} = u_{\lambda_k}\tau + v_{\lambda_k}, \quad u_{\lambda_k} := \operatorname*{arginf}_u \left\{ |u|_{BV} + \frac{\lambda_k}{\tau}\|v_{\lambda_{k-1}} - u\tau\|_{L^1} \right\}.$$

A telescoping sum of the refinement step $v_{\lambda_{k-1}} - v_{\lambda_k} = u_{\lambda_k}\tau$ yields the following *hierarchical* (BV, L^1) *decomposition* of f:

$$\begin{aligned} f &= u_{\lambda_0}\tau + v_{\lambda_0} \\ &= u_{\lambda_0}\tau + u_{\lambda_1}\tau + v_{\lambda_1} \\ &= \cdots \\ &= u_{\lambda_0}\tau + u_{\lambda_1}\tau + \cdots + u_{\lambda_N}\tau + v_{\lambda_N}. \end{aligned}$$

Remark 2.1. We note that in order to begin with only the main features of f, the initial scale needs to be small but not too small: to capture a nontrivial minimizer, λ_0 needs to be larger than a minimal threshold, which is quantified in terms of the dual $\|\cdot\|_*$-norm specified in (2.7) below [34, eq. (2.5)], [22]; i.e.,

$$\lambda_0 > \lambda_L := 1/\|\operatorname{sgn}(f)\|_*. \tag{2.3}$$

This yields a *hierarchical* (BV, L^1) *multiscale image decomposition*,

$$f = \sum_{k=k_0}^{N} u_{\lambda_k}\tau + v_{\lambda_N}, \tag{2.4}$$

where k_0 signals the initial scale λ_{k_0} to yield the first nontrivial minimizer u_{λ_0}. Consider the Nth step in the (BV, L^1) scheme,

$$u_{\lambda_N} = \operatorname*{arginf}_u \left(\int_\Omega |\nabla u| + \frac{\lambda_N}{\tau} \int_\Omega |v_{\lambda_{N-1}} - u\tau| \right).$$

The Euler–Lagrange equation associated with this minimizer reads (with the usual understanding of a regularized curvature term) as

$$\operatorname{sgn}\left(u_{\lambda_N}\tau - v_{\lambda_{N-1}}\right) = \frac{1}{\lambda_N} \operatorname{div}\left(\frac{\nabla u_{\lambda_N}}{|\nabla u_{\lambda_N}|}\right). \tag{2.5}$$

From (2.4) we have

$$v_{\lambda_{N-1}} = f - \sum_{k=k_0}^{N-1} u_{\lambda_k}\tau,$$

and using this above expression in (2.5) we find

$$\operatorname{sgn}\left(\sum_{k=k_0}^{N} u_{\lambda_k}\tau - f\right) = \frac{1}{\lambda_N}\operatorname{div}\left(\frac{\nabla u_{\lambda_N}}{|\nabla u_{\lambda_N}|}\right).$$

Letting $\tau \to 0$, we arrive at the following (BV, L^1) IDE:

$$\operatorname{sgn}\left(\int_{s=t_0}^{t} u(x,s)\,ds - f(x)\right) = \frac{1}{\lambda(t)}\operatorname{div}\left(\frac{\nabla u(x,t)}{|\nabla u(x,t)|}\right), \qquad \frac{\partial u}{\partial \mathbf{n}}\Big|_{\partial\Omega} = 0, \quad t \geq t_0, \tag{2.6}$$

subject to the initial condition $u(\cdot, t_0) = 0$. The scaling function $\lambda(t)$ is *any* monotone increasing function at our disposal. We discuss the role of this function in the next section.

2.2. On the scaling function $\lambda(t)$. It is argued in [22] that the dual norm,

$$\|w\|_* := \sup_{|\varphi|_{BV}\neq 0} \frac{\langle w, \varphi - \overline{\varphi}\rangle_{L^2}}{|\varphi|_{BV}}, \quad \overline{\varphi} = \frac{1}{|\Omega|}\int_\Omega \varphi(x)dx, \tag{2.7}$$

is a proper norm to measure texture (equivalently, one can characterize this dual norm as $\|w\|_* := \inf_g\{\|g\|_{L^\infty} : \operatorname{div} g = w - \overline{w}\}$). To understand the critical role of the scaling function $\lambda(t)$ in the IDE model (2.6) and its relationship with the "star-norm," we first prove the following useful lemma.

Lemma 2.1. *For $u \in BV$, let κ_u be the curvature term, $\kappa_u := \operatorname{div}\left(\frac{\nabla u}{|\nabla u|}\right)$, whenever $|\nabla u| \neq 0$. Then*

$$\|\kappa_u\|_* = 1.$$

Proof. For $\varphi \in BV$ we have the following:

$$|(\kappa_u, \varphi)_{L^2}| = \left|\left(\operatorname{div}\left(\frac{\nabla u}{|\nabla u|}\right), \varphi\right)_{L^2}\right| \leq |\varphi|_{BV}. \tag{2.8}$$

Thus, we have $\|\kappa_u\|_* \leq 1$. Letting $\varphi = u$ in (2.8), we obtain $|\langle \kappa_u, u\rangle_{L^2}| = |u|_{BV}$ and the result follows. ■

The following theorem is a direct consequence of this lemma.

Theorem 2.2. *Consider the IDE model* (2.6)

$$\operatorname{sgn}(U(x,t) - f(x)) = \frac{1}{\lambda(t)}\operatorname{div}\left(\frac{\nabla u(x,t)}{|\nabla u(x,t)|}\right), \qquad U(x,t) := \int_{t_0}^{t} u(x,s)\,ds.$$

Then, the size of the residual, $V(\cdot,t) := f - \int_{t_0}^{t} u(\cdot,s)\,ds$, is dictated by the scaling function $\lambda(t)$:

$$\left\|\operatorname{sgn}(V(\cdot,t))\right\|_* = \frac{1}{\lambda(t)}.$$

The "star-norm" measures oscillations [22]. Thus, the above equality could be interpreted as saying that the oscillations in the residual image $V(\cdot,t)$ are inversely proportional to the scaling function $\lambda(t)$. In particular, the oscillations in the residual function diminish in a precise manner controlled by increasing $\lambda(t)$.

Remark 2.2. Recall that t_0 in Theorem 2.2 is determined as the minimal *effective* scale in the sense that (2.3) holds; i.e., $\lambda(t_0) > \lambda_L = 1/\|\mathrm{sgn}(f)\|_*$. One does *not* need, however, to have a priori knowledge on the amount of "star-norm" oscillation in f, since $U(\cdot,t) \equiv 0$ as long as $\lambda(t)$ is underestimated, i.e., $\lambda < \lambda_L$. Eventually, the increasing $\lambda(t)$ will pick up the minimal effective scale.

2.3. L^1-energy decomposition. The increase of $\lambda(t)$ is directly linked to the "energy" of the underlying image. For the (BV, L^2)-based IDE (2.1), we proved in [32] the L^2-energy decomposition

$$\int_{s=t_0}^{t} \frac{1}{\lambda(s)} |u(\cdot,s)|_{BV}\, ds + \|V(\cdot,t)\|_{L^2}^2 = \|f\|_{L^2}^2.$$

For the (BV, L^1)-based IDE we have the corresponding L^1-energy decomposition.

Proposition 2.3. *Consider the IDE (2.6) and let $V(\cdot,t)$ denote the residual, $V(\cdot,t) := \int_{t_0}^{t} u(\cdot,s)\, ds - f$. Then we have the following L^1-energy decomposition:*

$$\int_{t_0}^{t} \frac{1}{\lambda(s)} |u(\cdot,s)|_{BV}\, ds + \|V(\cdot,t)\|_{L^1} = \|f\|_{L^1}.$$

Proof. We denote by $U(\cdot,t) := \int_{t_0}^{t} u(\cdot,s)\, ds$ the *cumulative image* so that the IDE (2.6) takes the form

$$\frac{1}{\lambda(t)} \operatorname{div}\left(\frac{\nabla u(x,t)}{|\nabla u(x,t)|}\right) = \mathrm{sgn}(U(x,t) - f(x)).$$

Integrating the above against $u \equiv \frac{d}{dt}(U - f)$, we get

$$\begin{aligned} -\frac{1}{\lambda(t)} |u(\cdot,t)|_{BV} &= \int_\Omega \mathrm{sgn}(U(\cdot,t) - f)\, \frac{d}{dt}\big(U(\cdot,t) - f\big)\, dx \\ &= \frac{d}{dt} \int_\Omega \mathrm{sgn}(U(\cdot,t) - f)\, (U(\cdot,t) - f)\, dx = \frac{d}{dt} \int_\Omega |U(\cdot,t) - f|\, dx, \end{aligned}$$

and the result follows by time integration. ■

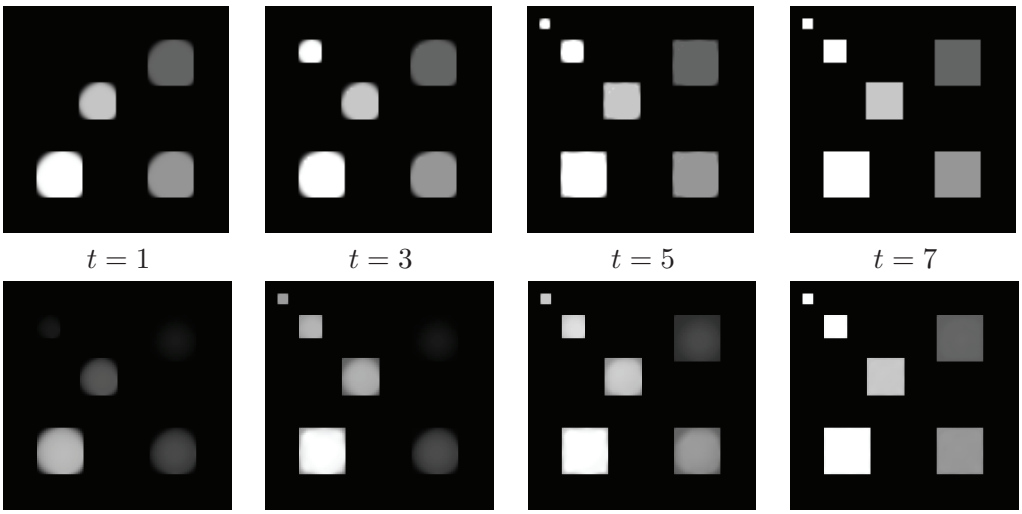
3. IDEs for images: Numerical results and extensions.

3.1. Comparing the (BV, L^1)- and (BV, L^2)-based IDEs. The multiscale image representation generated by the IDE (2.6) is shown in Figure 1. It is instructive to look at the image representation produced by the (BV, L^1)-based IDE (2.6) with the (BV, L^2)-based IDE (2.1). In the first row of Figure 2 we see the multiscale representation using the new (2.6). Note that the larger blocks appear before the small blocks, irrespective of their intensity levels. We observe that the appearance of these blocks is abrupt, as opposed to the results obtained with (2.1) shown in the second row of Figure 2, where these blocks appear gradually and the multiscale representation produced by (2.1) depends on the intensity levels rather than the size of the blocks. This feature makes (BV, L^1) IDE (2.6) more suitable for multiscale representation than (BV, L^2) IDE (2.1).

The difference between the (BV, L^1)-based IDE and the (BV, L^2)-based IDE is demonstrated clearly in Figure 3. We note that, for the same time, the (BV, L^1) IDE gives better

$t = 1$ $t = 10$ $t = 13$ $t = 25$

Figure 1. *The images* $U(t) = \int_0^t u(\cdot, s)\, ds$ *of the IDE* (2.6) *at* $t = 1, 10, 13, 25$. *Here,* $\lambda(t) = 0.01 \times 2^t$.



$t = 1$ $t = 3$ $t = 5$ $t = 7$

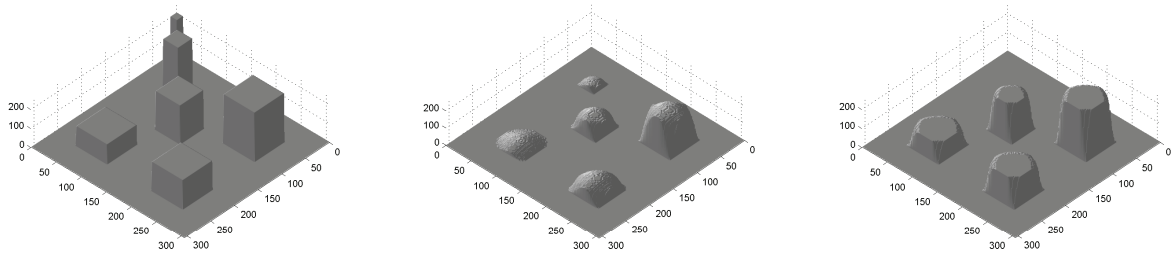
$t = 1$ $t = 4$ $t = 6$ $t = 10$

Figure 2. *First row: the images* $U(t) = \int_0^t u(\cdot, s)\, ds$ *of the* (BV, L^1)*-based IDE* (2.6) *at* $t = 1, 3, 5, 7$. *Here,* $\lambda(t) = 0.15 \times 2^t$. *Second row: the images* $U(t) = \int_0^t u(\cdot, s)\, ds$ *of the* (BV, L^2)*-based IDE* (2.1) *at* $t = 1, 4, 6, 10$. *Here,* $\lambda(t) = 0.1 \times 2^t$.

results at the edges than the (BV, L^2) IDE. The (BV, L^1) IDE is seen to be contrast invariant and depends only on the scale of the blocks, whereas the (BV, L^2) IDE is not contrast invariant, as expected.

3.2. IDE with filtered diffusion. Recall that one of the drawbacks in using the heat equation for denoising is that it results in an isotropic diffusion. The Perona–Malik model removes this drawback by introducing a diffusion controlling function that controls the diffusion near prominent edges in a given image. We propose a similar modification to our IDE model, seeking $u(x, t) : \Omega \times \mathbb{R}_+ \mapsto \mathbb{R}$ such that

$$\text{(3.1a)} \qquad \operatorname{sgn}\left(\int_{s=t_0}^{t} u(x, s)\, dx - f(x)\right) = \frac{g(|G_\sigma \star \nabla u(x, t)|)}{\lambda(t)} \operatorname{div}\left(\frac{\nabla u(x, t)}{|\nabla u(x, t)|}\right), \quad t > t_0,$$



(a) (b) (c)

Figure 3. *In* (a), (b), *and* (c) *we see the mesh views of the original image from Figure* 2, *the result* $U(t) = \int_0^t u(\cdot, s)\, ds$ *for the* (BV, L^2)*-based IDE* (2.1) *at* $t = 1$, *and the image* $U(t) = \int_0^t u(\cdot, s)\, ds$ *for the* (BV, L^1)*-based IDE* (2.6), *also at* $t = 1$, *respectively.*

subject to normal boundary conditions $\frac{\partial u}{\partial \mathbf{n}}|_{\partial\Omega} = 0$. As before, the IDE (3.1a) is activated after the minimal time scale, $t \geq t_0$, such that $u(\cdot, t_0)$ is the first slice of f to capture nontrivial features at that scale.

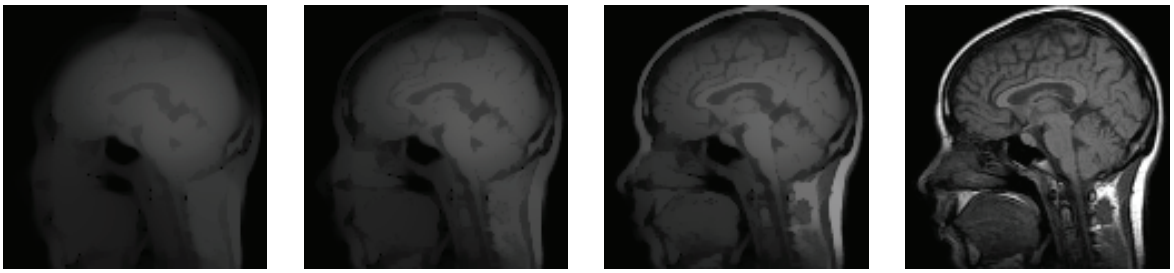
Similar to the Perona–Malik-type models, we can choose the prefactor function g so that it vanishes at infinity to control the diffusion at prominent edges in the image. Thus, the function g acts here as a high-pass filter which retains prominent edges in the image $\int_0^t u(x, s)\, ds$ without diffusing them. As choices for such a g-filter, Figure 5 displays the results of the modified IDE (3.1a) with

$$g(s) = \frac{1}{1 + (s/\beta)^2}. \tag{3.1b}$$

Here, the constant β determines the extent to which edges are preserved: for small β's, relevant edges are preserved, whereas for large β's, they are diffused. Detailed discussion of the numerical scheme for the filtered diffusion model (3.1) is given in section 4.

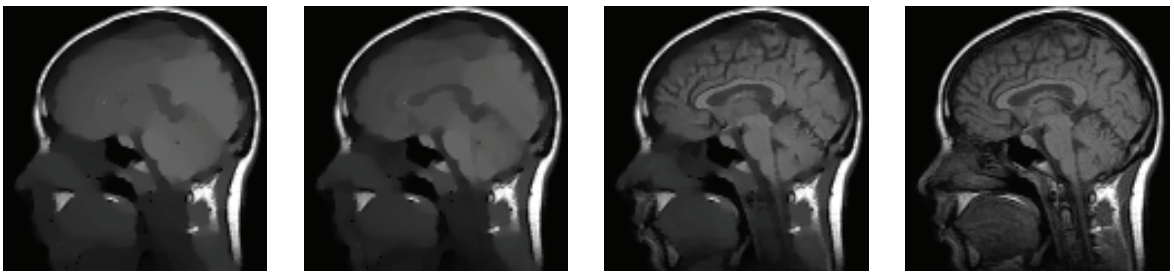
Comparing the results of the filtered IDE (3.1a) shown in Figure 5, we observe that edges, which are diffused by the basic IDE (2.6) as depicted in Figure 4, are preserved in Figure 5. As in the case of (2.1) (see [32]), this phenomenon is more apparent for smaller values of t due to the fact that as t increases, $U(\cdot, t)$ in both models approaches f and, consequently, suffers from less diffusion of the edges. The usefulness of the filtered diffusion IDE model becomes apparent when certain edges are required in the scale space for smaller values of t. For example, in Figure 4, the edges are blurred for smaller values of t with the standard IDE (2.6), but with the filtered diffusion IDE (3.1a) we retain relevant edges, as shown in Figure 5. We remark that this edge preserving filtered diffusion could be used in landmark-based hierarchical image registration algorithms [23, 28].

3.3. IDE with tangential smoothing. The approach of using the diffusion controlling function works well with natural images with moderate gradients. With other images, however, such as those which often arise in computer vision and industrial applications, the boundaries of their internal objects are marked with large, sharp gradients, for example, characteristic function χ_D, where $D \subset \Omega$. In such cases, we can choose to smooth only in the tangential



$t = 4$ $t = 4.2$ $t = 4.4$ $t = 6$

Figure 4. *The images* $U(t) = \int_0^t u(\cdot, s)\, ds$ *of the IDE* (2.6) *at* $t = 4, 4.2, 4.4, 6$. *Here,* $\lambda(t) = 0.1 \times 2^t$.



$t = 0.2$ $t = 3.8$ $t = 4$ $t = 6$

Figure 5. *The images* $U(t) = \int_0^t u(\cdot, s)\, ds$ *of the IDE* (3.1a) *at* $t = 0.2, 3.8, 4, 6$. *Here,* $\lambda(t) = 0.1 \times 2^t$.

direction to the boundaries of the objects, e.g., [5]. To this end, write $\Delta u := u_{TT} + u_{NN}$, where u_{TT} and u_{NN} are the tangential and normal diffusion components, i.e.,

$$u_{TT} = \Delta u - u_{NN} = |\nabla u| \operatorname{div}\left(\frac{\nabla u}{|\nabla u|}\right), \qquad u_{NN} = \left\langle \frac{\nabla u}{|\nabla u|}, \nabla^2 u \frac{\nabla u}{|\nabla u|} \right\rangle.$$

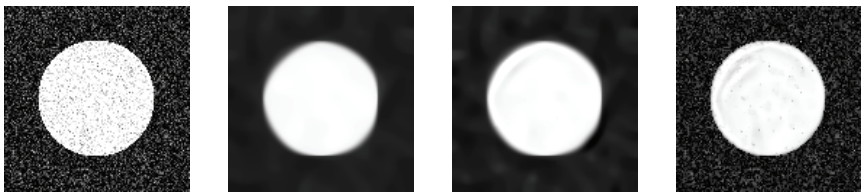
If we restrict the diffusion in our IDE model to tangential directions, we obtain modified IDEs with tangential smoothing,

$$\operatorname{sgn}\left(\int_{t_0}^t u(x,s)\, ds - f(x)\right) = \frac{1}{\lambda(t)} |\nabla u(x,t)| \operatorname{div}\left(\frac{\nabla u(x,t)}{|\nabla u(x,t)|}\right), \quad \frac{\partial u}{\partial \mathbf{n}}\Big|_{\partial\Omega} = 0, \tag{3.2}$$

and with tangential smoothing and filtering,

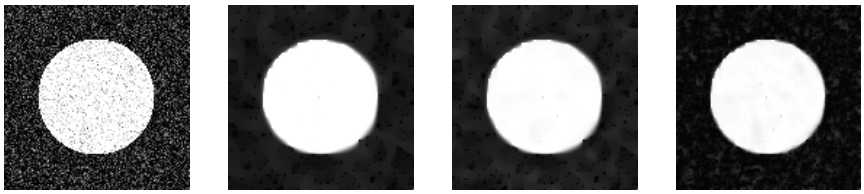
$$\operatorname{sgn}\left(\int_{t_0}^t u(x,s)\, ds - f(x)\right) = \frac{g(|G_\sigma \star \nabla u(x,t)|)}{\lambda(t)} |\nabla u(x,t)| \operatorname{div}\left(\frac{\nabla u(x,t)}{|\nabla u(x,t)|}\right). \tag{3.3}$$

As before, $u : \Omega \times \mathbb{R}_+ \mapsto \mathbb{R}$ evolves in inverse scale space using the conditions $u(\cdot, 0) \equiv 0$ and $\frac{\partial u}{\partial \mathbf{n}}|_{\partial\Omega} = 0$. Numerical experiments are shown in Figures 6 and 7. Compare the (BV, L^1) IDE results (2.6) shown in Figure 6 with the tangential smoothing (3.3) shown in Figure 7: the point here is that the tangential diffusion model preserves the edges, while denoising the rest of the image at a much faster rate than in the standard IDE model.



f at $t = 0$ $t = 1$ $t = 4$ $t = 18$

Figure 6. *A given noisy image f and the IDE images $\int_0^t u(\cdot, s)\, ds$ of (2.6) at $t = 1, 4, 18$. Here, the scaling function is $\lambda(t) = 0.0001 \times 2^t$. Most of the noise is present at scale $t = 18$.*



f at $t = 0$ $t = 1$ $t = 4$ $t = 18$

Figure 7. *A given noisy image f and the IDE images $\int_0^t u(\cdot, s)\, ds$ of (3.3) at $t = 1, 4, 18$. Here, the scaling function is $\lambda(t) = 0.0001 \times 2^t$. We observe that at $t = 18$ the noise is suppressed and at the same time the boundary of the circle is kept intact.*

4. Appendix: Numerical discretizations. In this appendix we describe the numerical implementation of (2.6) and its variants. First let us concentrate on the basic (BV, L^1) IDE model (2.6), rewritten here for convenience (for simplicity we set the initial scale at $\lambda(t_0)$ at $t_0 = 0$),

$$\operatorname{sgn}\left(\int_0^t u(x,s)\, ds - f(x)\right) = \frac{1}{\lambda(t)} \operatorname{div}\left(\frac{\nabla u(x,t)}{|\nabla u(x,t)|}\right). \tag{4.1}$$

In practice we approximate the BV-seminorm $|u|_{BV}$ with $\int_\Omega \sqrt{|\nabla u|^2 + \epsilon^2}$, which leads to the following IDE:

$$\operatorname{sgn}\left(\int_0^t u(x,s)\, ds - f(x)\right) = \frac{1}{\lambda(t)} \operatorname{div}\left(\frac{\nabla u(x,t)}{\sqrt{\epsilon^2 + |\nabla u(x,t)|^2}}\right). \tag{4.2}$$

In our numerical experiments we used the regularization parameter $\epsilon = 10^{-3}$. Then, as usual, $U(t) := \int_0^t u(x,s)\, ds$ is the exact solution. Let Δt be the time step, and let U^{n+1} denote the corresponding computed solution at $t^{n+1} = (n+1)\Delta t$:

$$U^{n+1} = U^n + W^{n+1}, \qquad W^{n+1} \equiv W_{i,j}^{n+1} := u_{i,j}^{n+1} \Delta t,$$

where $u_{i,j}^{n+1} \equiv u^{n+1}(ih, jh)$ is the approximate solution of the IDE at grid point (ih, jh). We introduce $\alpha^n := |U^n - f|$, and we compute the increment W^{n+1} by fixed-point iterations $\{\omega^k\}$.

With this, the IDE (4.2) is discretized at $t = t^{n+1}$:

$$
\begin{aligned}
&U^n_{i,j} + \omega^{k+1}_{i,j} = f_{i,j} \\
&+ \frac{\alpha^n}{\lambda^{(n+1)}h^2}\left[\frac{\omega^k_{i+1,j} - \omega^{k+1}_{i,j}}{\sqrt{\varepsilon^2 + (D_{+x}\omega^k_{i,j})^2 + (D_{0y}\omega^k_{i,j})^2}} - \frac{\omega^{k+1}_{i,j} - \omega^k_{i-1,j}}{\sqrt{\varepsilon^2 + (D_{-x}\omega^k_{i,j})^2 + (D_{0y}\omega^k_{i-1,j})^2}}\right] \\
&+ \frac{\alpha^n}{\lambda^{(n+1)}h^2}\left[\frac{\omega^k_{i,j+1} - \omega^{k+1}_{i,j}}{\sqrt{\varepsilon^2 + (D_{0x}\omega^k_{i,j})^2 + (D_{+y}\omega^k_{i,j})^2}} - \frac{\omega^{k+1}_{i,j} - \omega^k_{i,j-1}}{\sqrt{\varepsilon^2 + (D_{0x}\omega^k_{i,j-1})^2 + (D_{-y}\omega^k_{i,j})^2}}\right].
\end{aligned}
\tag{4.3}
$$

The nonlinear system (4.3) is solved using Jacobi iterations, which leads to the fixed-point iterations for computing ω^{k+1}:

$$
\omega^{k+1}_{i,j} = \frac{\lambda^{(n+1)}h^2(f_{i,j} - U^n_{i,j}) + \alpha^n(c_E\omega^k_{i+1,j} + c_W\omega^k_{i-1,j} + c_S\omega^k_{i,j+1} + c_N\omega^k_{i,j-1})}{\lambda^{(n+1)}h^2 + \alpha^n(c_E + c_W + c_S + c_N)}. \tag{4.4a}
$$

Here, $\lambda^{(n+1)} = \lambda(t^{n+1})$ are the discrete scaling parameters, and c_E, c_W, c_S, c_N are the discrete coefficients

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

In the computations above we set $h = 1$. To minimize the grid effects, we alternate the directions in which the above iterations are carried out, starting at the top-left corner position $(1, 1)$. Fixing $i = 1$, we vary $j = 1$ to j_{max} (east-south direction), initiating the next iteration at the top-right corner, and so on. The fixed-point iterations (4.4a) yield $\omega^k \overset{k\to\infty}{\longrightarrow} W^{n+1} \equiv u^{n+1}\Delta t$, and we can then update the computed image U,

$$
U^{n+1} = U^n + W^{n+1}. \tag{4.4b}
$$

Next, we consider the filtered IDE (3.1a), which is rewritten here for convenience as

$$
\left(\int_0^t u(x,s)\,ds - f(x)\right) = \frac{g(|G_\sigma \star \nabla u(x,t)|)}{\lambda(t)} \operatorname{div}\left(\frac{\nabla u(x,t)}{|\nabla u(x,t)|}\right).
$$

The only difference here is the additional diffusion controlling function $g(|G_\sigma \star \nabla u(x,t)|)$, where G_σ is the two-dimensional Gaussian smoothing with standard deviation σ. The function $g(s) = \frac{1}{1+(s/\beta)^2}$ with $\beta = 5$ is used in our numerical experiments. We approximate

$$
g(|G_\sigma \star \nabla u(x,t)|) \approx g\left(\left|G_\sigma \star \frac{\nabla \omega^n_{i,j}}{\Delta t}\right|\right),
$$

and the expression on the right enters into the right-hand side of (4.3). We end up with the same discrete IDE scheme (4.4) with $\lambda^{(n)} \mapsto \lambda^{(n)}/g(|G_\sigma \star \nabla \omega^n_{i,j}/\Delta t|)$.

REFERENCES

[1] W. K. Allard, *Total variation regularization for image denoising,* I. *Geometric theory*, SIAM J. Math. Anal., 39 (2007), pp. 1150–1190.

[2] W. K. Allard, *Total variation regularization for image denoising,* II. *Examples*, SIAM J. Imaging Sci., 1 (2008), pp. 400–417.

[3] W. K. Allard, *Total variation regularization for image denoising,* III. *Examples*, SIAM J. Imaging Sci., 2 (2009), pp. 532–568.

[4] S. Alliney, *Digital filters as absolute norm regularizers*, IEEE Trans. Signal Process., 40 (1992), pp. 1548–1562.

[5] L. Alvarez, P.-L. Lions, and J.-M. Morel, *Image selective smoothing and edge detection by nonlinear diffusion.* II, SIAM J. Numer. Anal., 29 (1992), pp. 845–866.

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

[7] F. Andreu-Vaillo, V. Casalles, and J. Mazón, *Parabolic Quasilinear Equations Minimizing Linear Growth Functionals*, Progr. Math. 223, Birkhäuser, Basel, Switzerland, 2004.

[8] G. Aubert and P. Kornprobst, *Mathematical Problems in Image Processing. Partial Differential Equations and the Calculus of Variations*, 2nd ed., Appl. Math. Sci. 147, Springer-Verlag, New York, 2006.

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

[10] A. Buades, T. Le, J.-M. Morel, and L. Vese, *Fast cartoon + texture image filters*, IEEE Trans. Image Process., 19 (2010), pp. 1978–1986.

[11] F. Catté, P.-L. Lions, J.-M. Morel, and T. Coll, *Image selective smoothing and edge detection by nonlinear diffusion*, SIAM J. Numer. Anal., 29 (1992), pp. 182–193.

[12] T. F. Chan and S. Esedoḡlu, *Aspects of total variation regularized L^1 function approximation*, SIAM J. Appl. Math., 65 (2005), pp. 1817–1837.

[13] T. F. Chan, S. Esedoḡlu, and M. Nikolova, *Algorithms for finding global minimizers of image segmentation and denoising models*, SIAM J. Appl. Math., 66 (2006), pp. 1632–1648.

[14] T. F. Chan and J. Shen, *Image Processing and Analysis: Variational, PDE, Wavelet, and Stochastic Methods*, SIAM, Philadelphia, 2005.

[15] T. F. Chan, J. Shen, and L. Vese, *Variational PDE models in image processing*, Notices Amer. Math. Soc., 50 (2003), pp. 14–26.

[16] V. Duval, J.-F. Aujol, and Y. Gousseau, *The TVL1 model: A geometric point of view*, Multiscale Model. Simul., 8 (2009), pp. 154–189.

[17] S. Esedoḡlu and S. J. Osher, *Decomposition of images by the anisotropic Rudin-Osher-Fatemi model*, Comm. Pure Appl. Math., 57 (2004), pp. 1609–1626.

[18] S. Geman and D. Geman, *Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images*, IEEE Trans. Pattern Anal. Mach. Intell., 6 (1984), pp. 721–741.

[19] P. Guidotti, *A new well-posed nonlinear nonlocal diffusion*, Nonlinear Anal., 72 (2010), pp. 4625–4637.

[20] S. Li, *Markov Random Field Modeling in Computer Vision*, Springer-Verlag, London, 1995.

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

[22] Y. Meyer, *Oscillating Patterns in Image Processing and Nonlinear Evolution Equations. The Fifteenth Dean Jacqueline B. Lewis Memorial Lectures*, Univ. Lecture Ser. 22, American Mathematical Society, Providence, RI, 2001.

[23] J. Modersitzki, *Numerical Methods for Image Registration*, Oxford University Press, New York, 2004.

[24] V. Morozov, *Regularization Methods for Ill-Posed Problems*, CRC Press, Boca Raton, FL, 1993. Translated from the 1987 Russian original.

[25] D. Mumford and J. Shah, *Boundary detection by minimizing functionals*, in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, CA, 1985, pp. 22–26.

[26] D. Mumford and J. Shah, *Optimal approximations by piecewise smooth functions and associated variational problems*, Comm. Pure Appl. Math., 42 (1989), pp. 577–685.

[27] M. Nikolova, *Minimizers of cost-functions involving nonsmooth data-fidelity terms. Application to the processing of outliers*, SIAM J. Numer. Anal., 40 (2002), pp. 965–994.

[28] D. Paquin, D. Levy, E. Schreibmann, and L. Xing, *Multiscale image registration*, Math. Biosci. Eng., 3 (2006), pp. 389–418.

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

[30] L. Rudin, S. Osher, and E. Fatemi, *Nonlinear total variation based noise removal algorithms*, Phys. D, 60 (1992), pp. 259–268.

[31] O. Scherzer, M. Grasmair, H. Grossauer, M. Haltmeier, and F. Lenzen, *Variational Methods in Imaging*, Appl. Math. Sci. 167, Springer-Verlag, New York, 2009.

[32] E. Tadmor and P. Athavale, *Multiscale image representation using novel integro-differential equations*, Inverse Probl. Imaging, 3 (2009), pp. 693–710.

[33] E. Tadmor, S. Nezzar, and L. Vese, *A multiscale image representation using hierarchical* (BV, L^2) *decompositions*, Multiscale Model. Simul., 2 (2004), pp. 554–579.

[34] E. Tadmor, S. Nezzar, and L. Vese, *Multiscale hierarchical decomposition of images with applications to deblurring, denoising and segmentation*, Commun. Math. Sci., 6 (2008), pp. 1–26.

[35] A. Tikhonov and V. Arsenin, *Solutions of Ill-Posed Problems*, Scripta Series in Mathematics, V. H. Winston & Sons, Washington, D.C., John Wiley & Sons, New York, 1977. Translated from the Russian.

[36] J. Weickert, *Anisotropic Diffusion in Image Processing*, Teubner-Verlag, Stuttgart, Germany, 1998.