Weighted and well-balanced anisotropic diffusion scheme for image denoising and restoration

V.B. Surya Prasath a,∗, D. Vorotnikov b

a Department of Computer Science, University of Missouri-Columbia, MO 65211-2060, USA
b CMUC, Department of Mathematics, University of Coimbra, 3001–454 Coimbra, Portugal

ARTICLE INFO

Article history:
Received 25 February 2012
Accepted 16 October 2013

ABSTRACT

Anisotropic diffusion is a key concept in digital image denoising and restoration. To improve the anisotropic diffusion based schemes and to avoid the well-known drawbacks such as edge blurring and ‘staircasing’ artifacts, in this paper, we consider a class of weighted anisotropic diffusion partial differential equations (PDEs). By considering an adaptive parameter within the usual divergence process, we retain the powerful denoising capability of anisotropic diffusion PDE without any oscillating artifacts. A well-balanced flow version of the proposed scheme is considered which adds an adaptive fidelity term to the usual diffusion term. The scheme is general, in the sense that, different diffusion coefficient functions can be utilized according to the need and imaging modality. To illustrate the advantage of the proposed methodology, we provide some examples, which are applied in restoring noisy synthetic and real digital images. A comparison study with other anisotropic diffusion based schemes highlight the superiority of the proposed scheme.

© 2013 Elsevier Ltd. All rights reserved.

1. Introduction

Image denoising is one of the foremost tasks in digital image processing pipeline. There exist various methodologies for removing noise in images and the area of image restoration and edge detection have been considered by many authors. Starting with the pioneering work of Perona and Malik [1], diffusion based partial differential equations (PDEs) are widely used in image noise removal and edge detection, see [2] for a review. Let u 0 be the noisy image which need to be restored by removing noise without removing salient structures in it. Mathematically, u 0 : Ω → ℝ represents a noisy version of a true image, and it is obtained by the following imaging process

u 0 = u + n, (1)

here we assume the noise process n is additive Gaussian noise with known mean and variance σ n. The image domain Ω ⊂ ℝ 2 is a bounded domain, usually a rectangle.

The Perona–Malik scheme (PM) can be written as a time dependent PDE, for x ∈ Ω

\[ \frac{\partial u(x, t)}{\partial t} = \text{div}(c(|\nabla u(x, t)|) \nabla u(x, t)) \] (2)

with u(x, 0) = u 0(x), i.e., the input noisy image is the initial datum, and the above PDE is run for a finite time T > 0 to obtain denoised image u(·, T). The choice of the diffusion function c : [0, ∞) → [0, ∞) is important in controlling the smoothing

∗ Corresponding author. Tel.: +1 573 882 8391; fax: +1 573 882 8318.
E-mail addresses: surya.iit@gmail.com, prasaths@missouri.edu (V.B. Surya Prasath), mitvorot@mat.uc.pt (D. Vorotnikov).

1468-1218/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.nonrwa.2013.10.004
and even enhancement of edges. In [1] the following two diffusion functions are considered

\[ c_{p1}(s) = \frac{1}{1 + (s/K)^2}, \quad c_{p2}(s) = \exp(-s/K^2) \]

(3)

where \( K > 0 \) is the contrast parameter. By such choices of nonlinear functions, PM PDE (2) avoids the over-smoothing property of the Heat equation. Good numerical results coupled with the fact that theoretically the PM PDE with such diffusion functions (3) is ill-posed [3], has generated an enormous interest in the mathematical image processing community, see [2] for a review. Moreover, an anisotropic PDE such as (2) can be considered as a gradient descent of a suitable energy functional [4–6]. The success of the anisotropic diffusion can be attributed to the fact that the PDE can be effectively discretized [7].

Though the PDE based schemes exhibit good denoising behavior, sometimes they can give artifacts such as staircasing or blocky regions. These drawbacks can occur due to various reasons, the primary one is the use of gradients to control diffusion. To avoid this, there have been efforts to use better control mechanisms for inhibiting diffusion in flat regions of the image. These techniques can be classified into three broad categories: (1) Use spatial or time regularization of the gradients [8–11] (2) Use a separate PDE to get better diffusion coefficients [12–15] (3) spatially adaptive diffusion coefficients [16–22]. Though the spatial regularization reduces the effect of noise in gradient computations, it can still give staircasing effects and can have poor localization of edges. In coupled PDE based schemes, apart from the expense of solving another PDE to get the edge map, it can inherit the problems of the original diffusion PDE. Spatially adaptive diffusion coefficient based scheme tries to balance these issues by providing a robust edge map for the diffusion to act upon. Here we consider an adaptive scheme which is based on this methodology. Recently, nonlocal diffusion operators were considered in [23–25] with corresponding wellposedness results. Another approach is to use higher order diffusion models [26] which can overcome staircasing artifacts.

Recently, Barcelos et al. [27,28] considered a well-balanced model inspired by the idea of mean curvature motion [29] and Nordstrom's biased PDE [30] approach. In this paper, we generalize such a model and consider weighted anisotropic diffusion schemes which incorporates adaptive information computed from the image at scale \( t \). Moreover, following Smolka [31] a modification of the image fidelity term is also done to improve the denoising capability of the PDE. Following [27], wellposedness of the proposed scheme is proved using the theory of viscosity solutions. Numerical examples in image denoising are given to highlight the proposed model.

The rest of the paper is organized as follows. Section 2 introduces the proposed weighted anisotropic diffusion scheme and a modification based on the well-balanced flow model of [27] is presented. In Section 3 we provide an analysis of the PDE model using viscosity solution theory of Crandall et al. [32]. Section 4 details the numerical aspects and shows comparison denoising results on noisy images. Finally, Section 5 concludes the paper.

2. Weighted well-balanced anisotropic diffusion

2.1. Well-balanced flow equation

The well-balanced flow (WBF) equation studied by Barcelos et al. [27,28] is based on total variation and can be generalized to the divergence process such as the Perona–Malik diffusivity:

\[ \frac{\partial u}{\partial t} = g(\nabla u) \text{div} (c(\nabla u) \nabla u) - \lambda (1 - g)(u - u_0) \]

(4)

where \( g(u \star \nabla G_\sigma) = (1 + |G_\sigma \star \nabla u|^2)^{-1} \) is known as the edge stopping function. The pre-smoothing with \( G_\sigma(x) = (2\pi \sigma)^{-1} \exp(-(x^2/2\sigma)) \), a Gaussian kernel of width \( \sigma \), is used to avoid noisy oscillations from the gradient computations. If the diffusion function is \( c(s) = s^{-1} \) (total variation (TV) [33]) then we recover the model studied in [27]. This TV diffusion function, in a sense, represents the borderline case from a class of decreasing diffusion functions. More faster decreasing functions can also be used, for example [34], \( c(s) = s^{-2} \), though wellposedness results for (4) cannot be obtained in these cases.

2.2. Weighted anisotropic diffusion

Fig. 1 shows the synthetic image used in our experiments. It consist of homogeneous regions separated by strong edges, gradual slope and a circle object with noisy oscillations. Fig. 1(b) and (c) show the smoothed gradient and diffusion function \( c_{p1} \) computed using the original image. These images show that edge map of the image is captured by the diffusion coefficient and highlights its importance in restoration. The diffusion coefficient \( c \) used in the PDE (4) can be influenced greatly by noise and gradient computations can be oscillatory. Fig. 1(e) and (f) show the smoothed gradient and diffusion function \( c_{p1} \) values respectively computed using the noisy image \( \nabla u_0 \). Clearly an edge map obtained in this way can lead to diffusion leakage and further iterations can propagate this oscillations which gives staircasing artifacts. Moreover, these gradient based diffusion functions give rise to edge pruning under evolution [35]. Hence, we need to use an adaptive measure which
Fig. 1. Diffusion PDE denoising depends on good edge maps and if the noise persist through the iterations, it leads to staircasing artifacts in the denoised version. (a) Original Kikis image used in the experiments. (b) Smoothed gradient $|G \ast \nabla u|$ of the original image, $\sigma = 2$ (black signifies higher values and white lower). (c) Edge map of the original image computed using the diffusion function $c_{pm}^1$ from (3) with $K = 20$. (d) Noisy image obtained by adding Gaussian noise of $\sigma_n = 30$ to the original image. (e) Smoothed gradient $|G \ast \nabla u_0|$ of the noisy image, $\sigma = 2$. (f) Edge map of the noisy image computed using the diffusion function $c_{pm}^1$ from (3) with $K = 20$.

can give a pixel-wise information to the diffusion function $c(x, |\nabla u|)$ in the divergence process. We propose the following class of functions for the diffusion PDE (6):

$$c(x, |\nabla u|) = \alpha(x) c_g(|\nabla u|).$$

(5)

Here, $\alpha$ is the adaptive parameter estimated at each pixel $x \in \Omega$. The function $c_g$ depends on the gradient image $|\nabla u|$ and can be chosen similar to (3). Note that, similar adaptive diffusion function studied in [16] is done for TV gradient function, i.e $c_g(|\nabla u|) = |\nabla u|^{-1}$. Further, the proposed scheme (6) is modified to include the balance term of [27], and thus provides a well-balanced flow in terms of noise removal and edge preservation. Thus, we consider the following general model (Nordstrom’s biased version [30]) based on PM PDE from Eq. (2):

$$\frac{\partial u}{\partial t} = g \text{div} \left(c(x, |\nabla u|) \nabla u\right) - \lambda(1 - g)(u - u_0)$$

(6)

where the parameter $\lambda$ balances the fidelity term and the usual divergence process. Here, we made the diffusion function $c(x, |\nabla u|)$ to depend on the spatial variable $x \in \Omega$ as well as the magnitude of the gradient $|\nabla u|$, which implies the introduction of inhomogeneity into the PDE. For this reason we call the PDE in Eq. (6) as weighted and well-balanced flow (WWBF) equation.

2.3. Choice of diffusion function, weight and other issues

The original Perona–Malik diffusion functions (3) represent two different behaviors with respect to the way the diffusion propagation is carried out. The $c_{pm1}$ prefers flat regions over edges and can inhibit higher gradients faster than the $c_{pm2}$ function. To make the presentation simple, throughout the article we use the $c_{pm1}$ diffusion function in all the PDEs. There exists various choices [17,19,21,22] for the weight function $\alpha$ in Eq. (6). The first choice is to use the classical inverse gradient approach [17], $\alpha(x) = (1 + K |\nabla u_0(x)|^2)^{-1}$, the other two choices are the slowed diffusion approach [36], and the Canny edge detector based parameter [22], $\alpha(x) = 1 - G \ast \text{Canny}(u(x, t))$. Fig. 2 illustrate the usage of adaptive diffusion coefficient against the traditional edge stopping function in front of the divergence term. Note that the edge stopping function $g$ acts as the ‘rate’ of the diffusion whereas the adaptive coefficient $\alpha$ controls the ‘amount’ of diffusion. In this sense, both the edge stopping function $g$ and the adaptive parameter $\alpha$ give complementary information for solving the denoising problem. Table 1 provides a succinct comparison of different weight functions from the literature with respect to image restoration. We utilize the inverse gradient function as the weight in the numerical experiments reported here and observed similar results with other adaptive parameter based functions.
Remark 1. The balancing term parameter $\lambda$ can also be made adaptive, see Gilboa et al. [39]. A spatially adaptive balance parameter $\lambda(x)$ can keep the textural component in the restored image while keeping the fidelity constraint.

Remark 2. Further generalizations of the well-balanced flow are also possible. For example, the diffusion coefficient can also be made to depend on the image $u$, i.e., $c(x, u, |\nabla u|)$. Such a generalization can lead to different diffusion flows and can be designed to influence the restoration process.

The wellposedness of the proposed PDE (6) can be proved using the viscosity solution theory and its discretized version satisfies the usual scale-space properties as well.

### Table 1
Comparison of different weight functions for image denoising and restoration. Note that $G_\sigma$, Gaussian kernel, $1_d$ is the indicator function for a set $A$, $\chi_1$ is a smooth edge indicator function, $\text{Var}_u^2(u)$ local variance of the image function $u$, for more details we refer to the corresponding references.

<table>
<thead>
<tr>
<th>Ref.</th>
<th>$\alpha(x)$</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>[21]</td>
<td>$(1 +</td>
<td>G_\sigma \ast \nabla u_0(x)</td>
<td>^2)^{-1}$</td>
</tr>
<tr>
<td>[36]</td>
<td>$1_{\text{deg}<em>1}(G</em>\sigma \ast \nabla u(x, t))/((1 +</td>
<td>G_\sigma \ast \nabla u</td>
<td>^2)/K^2)$</td>
</tr>
<tr>
<td>[19]</td>
<td>$1 + M_0 \chi_1$, $M_0 &gt; 0$ constant</td>
<td>Edge indication</td>
<td>Excessive blurring</td>
</tr>
<tr>
<td>[22]</td>
<td>$1_0 - G_\sigma \ast \text{Canny}(u(x, t))$</td>
<td>Retains multi-scale edges</td>
<td>Cannot handle high noise</td>
</tr>
<tr>
<td>[37]</td>
<td>$\exp(-\theta(\text{Var}_u^2(u(x, t)), \theta)/\delta)$</td>
<td>Contextual discontinuities</td>
<td>Stippled pattern artifacts</td>
</tr>
<tr>
<td>[38]</td>
<td>$1_x + 1_y \exp\left((-</td>
<td>G_\sigma \ast \nabla u(x, t)</td>
<td>/K)^2\right)$</td>
</tr>
</tbody>
</table>

Fig. 2. Edge stopping vs adaptive diffusion coefficients: (a) edge stopping function of the noisy image $(1 + |G_\sigma \ast \nabla u_0|^2)^{-1}$. (b) Inverse gradient based $c(x, |\nabla u|) = \alpha(x) |\nabla u|$, where $\alpha(x) = 1/(1 + K |G_\sigma \ast \nabla u|^2)$. (c) Slowed diffusion $c(x, |\nabla u|) = (G_\sigma \ast \nabla u)/(1 + |G_\sigma \ast \nabla u|^2/K^2)$. (d) Canny edge detector based $c(x, |\nabla u|) = \alpha(x) |\nabla u|$, where $\alpha(x) = 1 - G_\sigma \ast \text{Canny}(u(x, t))$.

Fig. 3. Effect of the balancing (fidelity term) on denoising the Kikis noisy image ($\sigma_n = 30$) using the PM PDE (2) with $c_{pm1}$. Each image shows the fidelity at different time stamps $t = 20, 40, 100, 200$ (black signifies higher values and white lower). Top row: classical fidelity $(1 - g) |u(x, t) - u_0(x)|$. Bottom row: adaptive fidelity $(1 - g) |u(x, t) - u(x, t - 1)|$.

Note that the fidelity term in Eq. (6) provides a complementary information using the noisy image $u(x, 0) = u_0(x)$. To further increase the denoising capability, we can make the classical image fidelity term $(u(x, t) - u(x, 0))$ in Eq. (6) to be adaptive, i.e., $(u(x, t) - u(x, t - 1))$, see Smolka [31]. Fig. 3 shows the effect of fidelity on denoising the noisy Kikis image (Fig. 1(d)) using the PM PDE (2) with diffusion coefficient $c_{pm1}$. Comparing the adaptive approach (Fig. 3, bottom row) with the classical fidelity (Fig. 3, top row), we can see that the adaptive process keeps edge details as the iteration increases. This, in turn, will aid the proposed WWBF PDE (6) to smooth the noisy image without destroying the salient edges.
3. Theoretical considerations

3.1. Preliminaries

Following [27,40], we study the proposed PDE
\[
\frac{\partial u}{\partial t} = g(G \ast \nabla u) \text{div} (c(x, |\nabla u|) \nabla u) - \lambda (1 - g(G \ast \nabla u))(u - u_0)
\]
using the viscosity solution theory of P. L. Lions et al. [32]. Here we admit generic convolution kernels $G$, which, in particular, can be the Gaussian kernels $G_\sigma$, and arbitrary spatial dimension $n > 1$.

Throughout this section, we employ Einstein’s summation convention. Let us first introduce two auxiliary functions depending on $x$ and $p$ from $\mathbb{R}^n$, a symmetric-matrix-valued one $a$ and a vector one $\chi$. We denote
\[
a_{ij}(x, p) = c(x, |p|)\delta_{ij} + c_y(x, |p|)\frac{p_ip_j}{|p|},
\]
\[
\chi_i(x, p) = \frac{\partial c(x, |p|)}{\partial x_i}.
\]

Here $\delta_{ij}$ is Kronecker’s delta, and $c_y$ is the partial derivative of $c(x, y)$ with respect to the second variable.

As usual, for the sake of simplification of the presentation, we consider the case of spatially periodic boundary conditions [40] for Eq. (7). Namely, we assume that there is an orthogonal basis $\{b_i\}$ in $\mathbb{R}^n$ so that
\[
u(\cdot, x) = u(\cdot, x + b_i), \quad x \in \mathbb{R}^n, \quad i = 1, \ldots, n.
\]
The problem is complemented with the initial condition
\[
u(0, x) = u_0(x),
\]
where $x \in \mathbb{R}^n$, and $u_0$ is Lipschitz and satisfies (10). Of course, $c$ (and thus $a$ and $\chi$) should also satisfy the same spatial periodicity restriction (with respect to $x$ but not to $y$ or $p$).

Let us introduce the following algebraic notion. Given a diagonal matrix $B$, let $\text{mod}(B)$ be the matrix whose entries are the absolute values of the entries of $B$. Furthermore, if $B$ is an arbitrary symmetric matrix, it can be represented as $Q^T DQ$, where $D$ is a diagonal matrix and $Q$ is an orthogonal one. Then we define $\text{mod}(B) = Q^T \text{mod}(D)Q$. It is straightforward to check that this definition does not depend on a particular choice of $D$ and $Q$. Observe also that $\text{mod}(B)$ is always positive-semidefinite, whereas $\text{mod}(B) = B$ when $B$ itself is positive-semidefinite.

We make the following assumptions:
\[
a, \chi \text{ are continuous, bounded, periodic in } x, \quad \text{continuously differentiable in } x,
\]
\[
\text{and their } x\text{-derivatives are uniformly (w.r.t. } p\text{) bounded,}
\]
\[
a_{ij}(x, p)\xi_i \xi_j \geq C \text{mod} \left( \frac{\partial a(x, p)}{\partial x_k} \right) \xi_i \xi_j, \quad k = 1, \ldots, n, \quad \xi, x, p \in \mathbb{R}^n,
\]
\[
g : \mathbb{R}^n \to \mathbb{R}, \quad 0 \leq g \leq 1, \quad \sqrt{g} \text{ is Lipschitz,}
\]
\[
G \in \mathcal{W}^1_2(\mathbb{R}^n) \quad \text{(note that we do not assume it to be space-periodic),}
\]
\[
\lambda \geq 0.
\]

Here and below $C$ stands for a generic positive constant, which can take different values in different lines.

Definition 1 (Viscosity Solution). A function $u$ from the space
\[
C[[0, T] \times \mathbb{R}^n] \cap L_\infty(0, T, W^1_\infty(\mathbb{R}^n))
\]
is a viscosity sub-/supersolution to (7), (10), (11) if, for any $\phi \in C^2([0, T] \times \mathbb{R}^n)$ and any point $(t_0, x_0) \in (0, T) \times \mathbb{R}^n$ of local maximum/minimum of the function $u - \phi$, one has
\[
\frac{\partial \phi(t_0, x_0)}{\partial t} - g ((u \ast \nabla G)(t_0, x_0)) \text{div} (c(x_0, |\nabla \phi(t_0, x_0)|) \nabla \phi(t_0, x_0))
\]
\[+
\lambda (1 - g ((u \ast \nabla G)(t_0, x_0)))(u(t_0, x_0) - u_0(x_0)) \leq 0 / \geq 0,
\]
and equalities (10), (11) hold in the classical sense. A viscosity solution is a function which is both a subsolution and a supersolution.
3.2. Main result

Theorem 1. (i) The problem (7), (10), (11) has a viscosity solution in class (18) for every positive T. Moreover,
\[ \inf_{\mathbb{R}^n} u_0 \leq u(t, x) \leq \sup_{\mathbb{R}^n} u_0. \] (20)

(ii) Assume that
\[ \left( \sqrt{a(x, p)} - \sqrt{a(z, p)} \right) \leq C|x - z|, \quad x, z, p \in \mathbb{R}^n. \] (21)

Then for any viscosity solutions u and v to (7), the following estimate holds
\[ \sup_{\mathbb{R}^n} |u(t, \cdot) - v(t, \cdot)| \leq \Phi(t) \sup_{\mathbb{R}^n} |u(0, \cdot) - v(0, \cdot)| \] (22)

with some non-decreasing continuous scalar function \( \Phi \) dependent on u and v.

Proof. Note that (20) is a direct consequence of the definition of viscosity solution: to get the second inequality, one can put \( \phi(t, x) = \delta t \), then, at the point \( (t_0, x_0) \), \( t_0 > 0 \), of the global maximum of \( u(t, x) - \delta t \), (19) gives \( \delta + \lambda(1 - g((u * \nabla G)(t_0, x_0))) \leq 0 \), whence \( u(t_0, x_0) < u_0(x_0) \), so we get a contradiction since \( u(t_0, x_0) \) is the global maximum point of \( u(t, x) - \delta t \); thus the function \( u(t, x) - \delta t \) attains its global maximum at \( t = 0 \), and it remains to let \( \delta \to +0 \); similarly one derives the first one.

Now, we establish a formal a priori estimate for \( \sup_{\mathbb{R}^n} |\nabla u| \). Observe that (7) is equivalent to
\[ \frac{\partial u}{\partial t} = g(u * \nabla G) \left[ a_{ij}(x, \nabla u)u_{x_i x_j} + \chi_i(x, \nabla u)u_{x_i} \right] - \lambda(1 - g(u * \nabla G))(u - u_0). \] (23)

Fix T. Differentiating (23) with respect to each \( x_k \), \( k = 1, \ldots, n \), multiplying by \( 2u_{x_k} \), and adding the results, we get
\[
\mathcal{L}(|\nabla u|^2) := \frac{\partial |\nabla u|^2}{\partial t} - ga_{ij}(x, \nabla u) \frac{\partial^2}{\partial x_i \partial x_j} |\nabla u|^2 - g \frac{\partial a_{ij}(x, \nabla u)}{\partial p_l} u_{x_j} \frac{\partial}{\partial x_i} |\nabla u|^2
= -2g a_{ij}(x, \nabla u)u_{x_k x_j} + 2g g(u * \nabla G) \cdot \left( u * \frac{\partial \nabla G}{\partial x_k} \right) a_{ij}(x, \nabla u)u_{x_j}u_{x_k}
+ 2g \frac{\partial a_{ij}(x, \nabla u)}{\partial x_k} u_{x_k x_j} u_{x_j} + 2g g(u * \nabla G) \cdot \left( u * \frac{\partial \nabla G}{\partial x_k} \right) \chi_i(x, \nabla u)u_{x_i}u_{x_k}
+ 2g \frac{\partial \chi_i(x, \nabla u)}{\partial x_k} u_{x_k} u_{x_i} - 2\lambda(1 - g)u_{x_k}u_{x_k} + 2\lambda(1 - g)(u_0)_{x_k}u_{x_k}
+ 2\lambda \nabla g(u * \nabla G) \cdot \left( u * \frac{\partial \nabla G}{\partial x_k} \right) (u - u_0)u_{x_k}. \] (24)

At this point, we need the following generalization of [40, Lemma 2.6].

Lemma 1. Let A and B be quadratic matrices of order n. Assume that B is symmetric, and there is a constant \( M \geq 0 \) such that
\[ MA_{ij} \xi_i \xi_j \geq \text{mod} (B)_{ij} \xi_i \xi_j, \quad \forall \xi \in \mathbb{R}^n. \] (25)

Then for any matrix U (of the same order but not necessarily symmetric) one has
\[ \text{Tr}^2(BU^T) \leq M \| B \| \text{Tr}(UAU^T), \] (26)

where \( \| \cdot \| \) denotes the operator norm of a matrix.

Proof. Formulas (25) and (26) are invariant with respect to orthogonal changes of bases. Thus, without loss of generality we may assume that B is already diagonalized by an orthogonal transform. Then
\[
\text{Tr}^2(BU^T) = (B_{ij}U_{ij})^2 \leq \| B \| (\text{mod} (B))_{ij} U_{ij}^2 \leq \| B \| (\text{mod} (B))_{ij} U_{ij} U_{ij}
= \| B \| (\text{mod} (B))_{ij} U_{ij} U_{ij} \leq M \| B \| |A_{ij}| U_{ij} U_{ij} = M \| B \| \text{Tr}(UAU^T). \]

38

This lemma gives opportunity to discharge the undesired influence of the second and the third terms in the right-hand side of (24). For the third one, due to the lemma, (14) and Cauchy’s inequality, we have

\[
\left| 2g \frac{\partial g}{\partial x_k} u_{x_k} u_{x_k} \right| \leq C g \left| u_{x_k} \right| \sqrt{a_{ij}(x, \nabla u) u_{x_k} u_{x_l} u_{x_k} u_{x_l}} \\
\leq g a_{ij}(x, \nabla u) u_{x_k} u_{x_k} + C |\nabla u|^2. \tag{27}
\]

Since our assumptions yield

\[
\left| u \ast \frac{\partial G}{\partial x_k} \right| \leq C, \tag{28}
\]

and

\[
|\nabla g| \leq C \sqrt{g}, \tag{29}
\]

an application of the lemma with \( A = B = a \) and \( M = 1 \) implies

\[
2Vg(u \ast \nabla G) \cdot \left( u \ast \frac{\partial \nabla G}{\partial x_k} \right) a_{ij}(x, \nabla u) u_{x_k} u_{x_l} \leq C \left| u_{x_k} \right| \sqrt{g a_{ij}(x, \nabla u) u_{x_k} u_{x_l} u_{x_k} u_{x_l}} \\
\leq g a_{ij}(x, \nabla u) u_{x_k} u_{x_k} + C |\nabla u|^2. \tag{30}
\]

The sum of the absolute values of the subsequent terms of the right-hand side of (24) does not exceed \( C(1 + |\nabla u|^2) \). Thus,

\[
\mathcal{L}(|\nabla u|^2) \leq C(1 + |\nabla u|^2), \tag{31}
\]

so

\[
\mathcal{L} \left( e^{-\mathcal{L}(t)} (1 + |\nabla u|^2) \right) \leq 0. \tag{32}
\]

From the weak maximum principle for the weakly parabolic operator \( \mathcal{L} \) one easily concludes that

\[
|\nabla u|^2 \leq C. \tag{33}
\]

Using (20) and (33), by means of the approach from [40] we can get the uniform Hölder estimate

\[
|u(t, x) - u(s, x)|^2 \leq C|t - s|. \tag{34}
\]

Then, following [42,40], we approximate our problem by well-posed ones in the sense of [43, Chapter 5]. Due to (20), (33) and (34), the solutions of these problems are uniformly bounded and equicontinuous on \([0, T] \times \mathbb{R}^n\). Then we can select a uniformly converging sequence of approximate solutions, and pass to the limit in the viscosity sense using the general consistency/stability results from [32]. The uniqueness of solutions follows from the stability estimate (22). This bound may be shown by revisiting the proof of a similar bound in [42,40]. We only point out that the matrix \( \Gamma^* \) [40, p. 159] is replaced by

\[
\Gamma^* = \left( \frac{g_1 A_1}{\sqrt{g_1 g_2} \sqrt{A_2} \sqrt{A_1}} \frac{\sqrt{g_1 g_2} \sqrt{A_1} \sqrt{A_2}}{g_2 A_2} \right), \tag{35}
\]

where

\[
A_1 = a \left( x_0, \frac{|x_0 - y_0|^2 (x_0 - y_0)}{\varepsilon} \right), \quad A_2 = a \left( y_0, \frac{|x_0 - y_0|^2 (x_0 - y_0)}{\varepsilon} \right),
\]

and the notation within is taken from [40]. Note that the \( 2n \times 2n \)-matrix \( \Gamma^* \) is symmetric and positive-semidefinite. □

4. Numerical results

4.1. Comparison with other schemes

The proposed scheme is compared with related diffusion based denoising schemes from the literature. To make a fair comparison we utilize the same diffusion function \( c_{pm1} \) from [3] in all the compared schemes and the contrast parameter \( K \) is fixed using the original criteria given in [1], see [44,45] for other choices. Moreover, the edge stopping function \( g(\xi) = (1 + |\xi|)^{-1} \) is fixed wherever applicable and the classical fidelity term is utilized unless otherwise stated.
(a) Perona and Malik [1] — Anisotropic Diffusion (AD) Eq. (2) with $c_{pm}$ in (3):
\[
\frac{\partial u}{\partial t} = \text{div} \left( \frac{\nabla u}{1 + |\nabla u|^2 / K^2} \right)
\]

(b) Catté et al. [8] — Smoothed Gradient based anisotropic diffusion (SG) with $c_{pm}$ in (3):
\[
\frac{\partial u}{\partial t} = \text{div} \left( \frac{\nabla u}{1 + |\nabla G_\sigma * u|^2 / K^2} \right)
\]

(c) Rudin et al. [33] — Total Variation (TV) (2) with $c(s) = (\epsilon + s^2)^{-1/2}$, $\epsilon = 10^{-6}$:
\[
\frac{\partial u}{\partial t} = \text{div} \left( \frac{\nabla u}{\sqrt{\epsilon + |\nabla u|^2}} \right)
\]

(d) El Falah and Ford [29] — Mean Curvature Motion (MCM), Eq. (4) with $\lambda = 0$:
\[
\frac{\partial u}{\partial t} = \frac{1}{1 + |\nabla u|^2} \text{div} \left( \frac{\nabla u}{1 + |\nabla u|^2 / K^2} \right)
\]

(e) Barcelos et al. [27] — Well-Balanced Flow (WBF):
\[
\frac{\partial u}{\partial t} = g(|\nabla G_\sigma * u|) \text{div} \left( \frac{\nabla u}{1 + |\nabla u|^2 / K^2} \right) - \lambda (1 - g(|\nabla G_\sigma * u|))(u - u_0)
\]

(f) Shi and Chang [9] — Modified Smoothed Gradient based anisotropic diffusion (MSG):
\[
\frac{\partial u}{\partial t} = |\nabla G_\sigma * u| \text{div} \left( \frac{\nabla G_\sigma * u}{|\nabla G_\sigma * u|} \right) - |\nabla G_\sigma * u| \lambda (G_\sigma * u - u_0)
\]

Further, similar adaptive schemes which utilize different diffusion coefficient functions are also compared.

(a) Weickert [46] — Edge Enhancing Diffusion (EED):
PM PDE (2) with the diffusion function:
\[
c(|\nabla u|) = \begin{cases} 
\exp(-0.234 |\nabla u|) & \text{if } |\nabla u| \geq K \\
0 & \text{if } |\nabla u| < K
\end{cases}
\]

(b) Weickert [47] — Coherence Enhancing Diffusion (CED):
PM PDE (2) with the diffusion function constructed using the structure tensor, see [47] for more details. The eigenvalues of $\mathcal{D}$ are chosen as, for $\mu_1, \mu_2$ eigenvalues of the structure tensor, $\alpha \in (0, 1)$, $C > 0$: $\lambda_1 = \alpha$, and
\[
\lambda_2 = \begin{cases} 
\alpha & \text{if } \mu_1 = \mu_2 \\
\alpha + (1 - \alpha) \exp \left( -\frac{C}{(\mu_1 - \mu_2)^2} \right) & \text{else}
\end{cases}
\]

(c) Kačur and Mikula [48,36] — Slowed Anisotropic Diffusion (SAD):
\[
\frac{\partial u}{\partial t} = \text{div} \left( \frac{\nabla G_\sigma * u}{1 + |\nabla G_\sigma * u|^2 / K^2} \nabla \beta(x, u) \right)
\]
with $\beta(x, u) = 0$ for $u \in [0, 0.5]$ and $\beta(x, u) = u$ for $u \in (0.5, 1]$.

(d) Strong [16] — Adaptive TV (ATV):
\[
\frac{\partial u}{\partial t} = \text{div} \left( \frac{\alpha(x) \nabla u}{\sqrt{\epsilon + |\nabla u|^2}} \right)
\]
with $\alpha(x) = (1 + |\nabla u_0|)^{-1}$, $\epsilon = 10^{-6}$

(e) Kusnezow et al. [19] — Adaptive Linear Diffusion (ALD):
\[
\frac{\partial u}{\partial t} = \alpha \text{div} (\nabla u)
\]
with $\alpha = (1 + M_e \chi_c)$, $M_e \gg 0$ constant and $\chi_c$ is a smooth edge indicator function.

(f) Prasath and Singh [22] — Edge detector based Anisotropic Diffusion (EAD):
\[
\frac{\partial u}{\partial t} = \text{div} \left( \frac{\alpha(x) \nabla u}{1 + |\nabla u|^2 / K^2} \right)
\]
with $\alpha(x) = 1 - G_\sigma * C(u(x, t))$, $C$-Canny edge detector output.
In the comparison results, apart from using the proposed adaptive fidelity term based WWBF (see Eq. (6) and Section 2.3),
\[
\frac{\partial u(x, t)}{\partial t} = g \text{ div } (c(x, |\nabla u|) \nabla u(x, t)) - \lambda (1 - g) (u(x, t) - u(x, t - 1)),
\]
we also utilize the weighted Linear Diffusion (WLD) — using the proposed weight in a linear diffusion framework,
\[
\frac{\partial u}{\partial t} = g \text{ div } (\alpha(x) \nabla u) - \lambda (1 - g) (u(x, t) - u(x, t - 1)).
\]

### 4.2. Implementation details

The additive operator splitting (AOS) scheme which is proven to be effective in diffusion PDE based image processing [7] is used to implement the schemes. The images were scaled to the interval [0, 1]. It can be described briefly as follows: In 1-D with matrix–vector notation, the iterative scheme is,
\[
U^{i+1} = \left[1 - \tau A(U^{i})\right]^{-1} U^{i},
\]
where \( \tau \) is the time step, \( A(U^{i}) = [a_{ij}(U^{i})] \), and
\[
a_{ij}(U^{i}) := \begin{cases} 
\frac{\gamma_{i}^{j} + \gamma_{j}^{i}}{2h^2} & j \in \mathcal{N}_{i} \\
- \sum_{k \in \mathcal{N}_{i}} \frac{\gamma_{i}^{k} + \gamma_{k}^{i}}{2h^2} & j = i \\
0 & \text{otherwise}
\end{cases}
\]
with \( \gamma_{i} = \alpha_{i}g_{i} \) and \( h \) discretization step size. For n-D images the semi-implicit scheme is written as
\[
U^{i+1} = \left[1 - \tau \sum_{l=1}^{n} A_{l}(U^{i})\right]^{-1} U^{i}. \tag{36}
\]

The matrix \( A_{l} = (a_{ij})_{ij} \) corresponds to derivatives along the \( l \)-th coordinate axis.

**Remark 3.** The spatial step size \( h = 1 \) is fixed as the pixel grid has the natural spacing of size one. Further the time step \( \tau = 0.2 \), pre-smoothing parameter \( \sigma = 1 \), and fidelity parameter \( \lambda = 1 \) are fixed for all the experiments reported here.

**Remark 4.** Under the AOS type discretization (36), the proposed WWBF scheme (6) satisfies the usual scale space properties, see [7] for more details. Moreover the maximum–minimum principle also holds, see Theorem 1.

### 4.3. Visual comparison

Fig. 4 shows the comparison of non-adaptive diffusion schemes based on restoration results for a noisy (Gaussian noise, \( \sigma_{n} = 25 \)) Lena gray-scale image. In each pair, left image shows the 156 \times 156 crop of the restored image and the right image shows the contour view to highlight the movement of level-sets under different schemes. Note that, the proposed approach gives better result even with linear diffusion, see Fig. 4(g) which corresponds to WLD scheme result. As can be seen by comparing the contour maps of each scheme, the proposed scheme's result in Fig. 4(h) gives better result in terms denoising as well as staircasing artifact free restoration.

To compare the adaptive diffusion schemes in a fair manner we utilize a test image synthetically generated consist of a slope, strong edges and a circle with oscillations. Fig. 5 shows the comparison results for the noisy Kikis image (\( \sigma_{n} = 30 \) is added to the original image, see Fig. 1(d)) by different adaptive diffusion schemes. The Perona–Malik, TV based schemes such as EED, CED, SAD, ATV inherit the original staircasing artifacts whereas WWBF performs better than other schemes in terms of edge preservation without oscillations, see Fig. 5(h).

Finally, to show the effect of the adaptive fidelity term in different adaptive schemes we perform experiments on a synthetic Circles gray-scale image which has multiple circular regions with different piecewise constant regions. Fig. 6 shows the comparison of the adaptive schemes SAD, ATV, ALD, and EAD with the same adaptive fidelity chosen as in our WWBF scheme, i.e., \( u(x, t) - u(x, t - 1) \). As can be seen the WWBF scheme preserves edges without any blocky artifacts. Moreover, the adaptive fidelity term helps capturing circular edges by using balanced adaptive diffusion. Supplementary MATLAB .fig files are provided for 3D visualizations of resultant images shown on the left of each sub-figure (see Appendix A).

**Remark 5.** Other non-adaptive diffusion schemes such as AD, SG, TV, MCM, WBF, MSG and directional diffusion models such as EED, CED do not utilize an adaptive weight as in our case (see Eq. (5)). Moreover, the adaptive data fidelity term did not provide any visually improved denoising results for these schemes, hence we omit the images in Fig. 6 for brevity.
4.4. Quantitative comparison and discussion

To compare the schemes quantitatively we utilize two commonly used error metrics in the image denoising literature, one is the classical peak signal to noise ratio (PSNR) [2], and the other is the mean structural similarity measure (MSSIM) [49]:

1. **PSNR** is given in decibels (dB). A difference of 0.5 dB can be identified visually. Higher PSNR value indicates optimum denoising capability.

\[
\text{PSNR}(u) := 20 \times \log_{10} \left( \frac{u_{\text{max}}}{\sqrt{\text{MSE}}} \right) \text{ dB}
\]

where \( \text{MSE} = (mn)^{-1} \sum \sum (u - u_0) \), \( m \times n \) denotes the image size, \( u_{\text{max}} \) denotes the maximum value, for example in 8-bit images \( u_{\text{max}} = 255 \).

2. **MSSIM** index is in the range \([0, 1]\). The MSSIM value near one implies the optimal denoising capability of the scheme [49] and is mean value of the SSIM metric. The SSIM is calculated between two windows \( \omega_1 \) and \( \omega_2 \) of common size.
Fig. 5. Adaptive schemes comparison results on Kikis $128 \times 128$ synthetic image, (in each sub-figure, the right image shows the surface form of the left image): (a) EED [46]; (b) CED [47]; (c) SAD [48]; (d) ATV [16]; (e) ALD [19]; (f) EAD [22]; (g) original image and its surface form given for comparison; (h) proposed scheme with nonlinear diffusion (WWBF).

$$N \times N,$$

$$\text{SSIM}(\omega_1, \omega_2) = \frac{(2\mu_{\omega_1}\mu_{\omega_2} + c_1)(2\sigma_{\omega_1\omega_2} + c_2)}{(\mu^2_{\omega_1} + \mu^2_{\omega_2} + c_1)(\sigma^2_{\omega_1} + \sigma^2_{\omega_2} + c_2)}$$

where $\mu_{\omega_1}$ the average of $\omega_1$, $\sigma^2_{\omega_1}$ the variance of $\omega_1$, $\sigma_{\omega_1\omega_2}$ the covariance, $c_1$, $c_2$ stabilization parameters, see [49] for more details.

Table 2 shows the comparison results using these two metrics for all schemes without data adaptive fidelity term. Corresponding PSNR and MSSIM values are given for each of the schemes and clearly our scheme performs better than or on par with other diffusion schemes in general. We also include comparison results with corresponding data adaptive fidelity term described in Section 2.3. As can be noted, the proposed scheme performs well for a variety of images (taken from the standard test images USC-SIPI database) for both data fidelity versions. Note that the PSNR values are closer together when adaptive fidelity is used (SAD, ATV, ALD, EAD, and our WWBF) in Table 2, but MSSIM values indicate a better performance of the proposed approach. Thus, the proposed adaptive WWBF flow preserves salient structures (edges) when compared with other nonlinear heat diffusion flows. The Baboon image consists of texture parts and hence the proposed
Fig. 6. Adaptive schemes comparison results on Circles $120 \times 120$ synthetic image, (in each sub-figure, right image shows the resultant and left image, the adaptive fidelity term at the final iteration in surface format): (a) SAD [48]; (b) ATV [16]; (c) ALD [19]; (d) EAD [22]; (e) original image and its edge map given in surface format for comparison; note that artifacts are due to jpeg compression which appear near edges; (f) proposed scheme with nonlinear diffusion (WWBF); supplementary MATLAB .fig files are provided for 3D visualizations of resultant images shown in the Appendix A.

Table 2

<table>
<thead>
<tr>
<th>Scheme</th>
<th>Ref.</th>
<th>Kikis</th>
<th>Lena</th>
<th>House</th>
<th>Peppers</th>
<th>Baboon</th>
</tr>
</thead>
<tbody>
<tr>
<td>Noisy</td>
<td></td>
<td>18.56/0.1683</td>
<td>20.14/0.3866</td>
<td>20.14/0.2732</td>
<td>20.14/0.3426</td>
<td>20.14/0.4643</td>
</tr>
<tr>
<td>AD [1]</td>
<td>33.32/0.9081</td>
<td>26.32/0.7752</td>
<td>28.87/0.8300</td>
<td>27.37/0.8170</td>
<td>23.48/0.4687</td>
<td></td>
</tr>
<tr>
<td>SG [8]</td>
<td>29.09/0.9036</td>
<td>23.26/0.6708</td>
<td>24.94/0.7657</td>
<td>23.09/0.7291</td>
<td>22.35/0.3653</td>
<td></td>
</tr>
<tr>
<td>TV [33]</td>
<td>33.47/0.9414</td>
<td>27.05/0.7951</td>
<td>30.18/0.8520</td>
<td>28.30/0.8389</td>
<td>23.61/0.4899</td>
<td></td>
</tr>
<tr>
<td>MCM [29]</td>
<td>30.87/0.9238</td>
<td>23.97/0.6943</td>
<td>25.89/0.7855</td>
<td>24.02/0.7501</td>
<td>22.44/0.3723</td>
<td></td>
</tr>
<tr>
<td>WBF [27]</td>
<td>33.19/0.9111</td>
<td>26.46/0.7827</td>
<td>28.94/0.8286</td>
<td>27.63/0.8289</td>
<td>23.54/0.4802</td>
<td></td>
</tr>
<tr>
<td>MSG [9]</td>
<td>33.23/0.9273</td>
<td>26.53/0.7826</td>
<td>29.30/0.8370</td>
<td>27.31/0.8327</td>
<td>23.34/0.4627</td>
<td></td>
</tr>
<tr>
<td>EED [46]</td>
<td>35.23/0.9530</td>
<td>27.23/0.7980</td>
<td>30.68/0.8554</td>
<td>28.44/0.8435</td>
<td>23.47/0.4685</td>
<td></td>
</tr>
<tr>
<td>CED [47]</td>
<td>30.87/0.9238</td>
<td>23.97/0.6943</td>
<td>25.89/0.7855</td>
<td>24.02/0.7501</td>
<td>22.44/0.3723</td>
<td></td>
</tr>
<tr>
<td>SAD [48]</td>
<td>34.57/0.9593</td>
<td>25.85/0.7559</td>
<td>29.18/0.8375</td>
<td>26.93/0.8030</td>
<td>22.90/0.4133</td>
<td></td>
</tr>
<tr>
<td>ATV [16]</td>
<td>33.68/0.9435</td>
<td>27.26/0.7972</td>
<td>30.39/0.8541</td>
<td>28.51/0.8410</td>
<td>23.82/0.4920</td>
<td></td>
</tr>
<tr>
<td>ALD [19]</td>
<td>28.48/0.9378</td>
<td>20.70/0.5965</td>
<td>23.23/0.7346</td>
<td>20.58/0.6248</td>
<td>20.84/0.3105</td>
<td></td>
</tr>
<tr>
<td>EAD [22]</td>
<td>34.24/0.9600</td>
<td>24.88/0.7247</td>
<td>28.17/0.8221</td>
<td>25.72/0.7719</td>
<td>22.47/0.3762</td>
<td></td>
</tr>
<tr>
<td>WLD 32.81/0.9423</td>
<td>23.87/0.7126</td>
<td>27.03/0.8088</td>
<td>23.91/0.7306</td>
<td>20.73/0.3557</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WWBF 37.00/0.9499</td>
<td>27.12/0.7815</td>
<td>30.92/0.8584</td>
<td>28.27/0.8109</td>
<td>22.98/0.4417</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

WWBF scheme cannot obtain optimal PSNR/MSSIM values. To alleviate this a spatially adaptive fidelity parameter $\lambda = \lambda(x)$ can be incorporated, see Section 2.3. Following [28] automatic selection of parameters is one of the current research being
carried out. Moreover, the image restoration model studied here can be used in other image processing algorithms such as inpainting [50,51] and edge detection [27] as well.

5. Conclusions

Well-balanced flow is based on a nonlinear diffusion PDE which is utilized in image noise removal and edge detection successfully. In this paper, a new variant of the flow is considered by using weights in the divergence diffusion process. This improves the denoising capabilities as well as the multi-scale detail preservation of the corresponding PDE. Numerical experiments on noisy images shows the proposed scheme’s performs well on a variety of images. Extensive experiments indicate the improvements over other classical diffusion and adaptive diffusion schemes.

Acknowledgments

The authors would like to thank the anonymous referees for their constructive comments which improved the paper considerably. Part of this work was done while the first author was visiting Institute for Pure and Applied Mathematics (IPAM), University of California Los Angeles (UCLA), USA. The first author thanks the IPAM institute for their great hospitality and support during the visit. The second author was partially supported by CMUC and FCT (Portugal), through European program COMPETE/FEDER under the projects PEst-C/MAT/UI0324/2013, PTDC/MAT/098060/2008, UTA-CMU/MAT/0007/2009, and PTDC/MAT–CAL/0749/2012.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at http://dx.doi.org/10.1016/j.nonrwa.2013.10.004.

References


