Available at
www.ElsevierComputerScience.com
POWERED BY SCIENCE DIRECT®

Image and Vision Computing 22 (2004) 73–81

www.elsevier.com/locate/imavis

A common framework for nonlinear diffusion, adaptive smoothing, bilateral filtering and mean shift

Danny Barash[a,*], Dorin Comaniciu[b]

[a]*Department of Chemistry and Courant Institute of Mathematical Sciences, New York University and Howard Hughes Medical Institute, 31 Washington Place, Main 1021, New York, NY 10003, USA*

[b]*Real-Time Vision and Modeling Department, Siemens Corporate Research, 755 College Road East, Princeton, NJ 08540, USA*

Received in revised form 29 July 2003

Abstract

In this paper, a common framework is outlined for nonlinear diffusion, adaptive smoothing, bilateral filtering and mean shift procedure. Previously, the relationship between bilateral filtering and the nonlinear diffusion equation was explored by using a consistent adaptive smoothing formulation. However, both nonlinear diffusion and adaptive smoothing were treated as local processes applying a 3×3 window at each iteration. Here, these two approaches are extended to an arbitrary window, showing their equivalence and stressing the importance of using large windows for edge-preserving smoothing. Subsequently, it follows that bilateral filtering is a particular choice of weights in the extended diffusion process that is obtained from geometrical considerations. We then show that kernel density estimation applied in the joint spatial–range domain yields a powerful processing paradigm—the mean shift procedure, related to bilateral filtering but having additional flexibility. This establishes an attractive relationship between the theory of statistics and that of diffusion and energy minimization. We experimentally compare the discussed methods and give insights on their performance.
© 2003 Elsevier B.V. All rights reserved.

Keywords: Nonlinear diffusion; Adaptive smoothing; Bilateral filtering; Mean shift procedure

1. Introduction

Nonlinear operations are becoming increasingly important in visual processing applications. Since they are substantially more difficult to analyze, formulate and predict compared to linear operations, various innovative approaches have been proposed independently for low-level computer vision tasks. The integration of several approaches that rely on different mathematical tools (e.g. functional minimization, nonlinear PDEs, statistics and data analysis) is essential for obtaining high-quality results in real-life applications.

This paper concentrates on edge-preserving smoothing. It extends previous work [1,2] on the relationship between nonlinear diffusion [19,20,26], adaptive smoothing [21], and bilateral filtering [24] to establish a connection to the mean shift procedure [8,9] in the joint spatial–range domain. Both nonlinear diffusion and adaptive smoothing are generalized to encompass large neighborhoods, while the bilateral filtering serves as a link between the *extended nonlinear diffusion* (i.e. nonlinear diffusion on extended neighborhoods) and mean shift filtering.

The paper is divided as follows. Section 2 emphasizes the importance of extended neighborhoods in edge-preserving smoothing by analyzing smoothing on 1D 3-neighborhood, smoothing on 1D 5-neighborhood, and adaptive smoothing on 1D 5-neighborhood. This leads to the formulation in Section 3 of the extended nonlinear diffusion on 2D $(2S+1 \times 2S+1)$-neighborhood. In Section 4, it is shown that a specific choice of weights in the extended nonlinear diffusion, that is based on geometrical considerations, leads to bilateral filtering. By defining kernel density estimation in the spatial–range domain, we derive in Section 5 the mean shift procedure for filtering and show its extended flexibility over bilateral filtering. In Section 6 experiments and comparisons are presented, while in Section 7, conclusions

* Corresponding author. Address: Genome Diversity Center, Institute of Evolution, Haifa University, Mount Carmel 31905, Isreal.

E-mail addresses: dbarash@research.haifa.ac.il (D. Barash), comanici@scr.siemens.com (D. Comaniciu).

0262-8856/$ - see front matter © 2003 Elsevier B.V. All rights reserved.
doi:10.1016/j.imavis.2003.08.005

are drawn based on the common framework that unifies several fundamental approaches for low-level vision.

2. Importance of extended neighborhood

The extension of gradient based, edge-preserving smoothing to include information from non-nearest neighboring pixels is natural and has been considered before in various contexts (e.g. Refs. [4,24,27]). Here, we start from Saint-Marc–Chen–Medioni's adaptive smoothing [21] that was reformulated in Ref. [2] for consistency with the diffusion equation, and extend the approach from the original 3×3 window to a window of arbitrary size.

The adaptive smoothing approach is fundamental and intuitive. Given an image $I^{(t)}(\vec{x})$, where $\vec{x} = (x_1, x_2)$ denotes space coordinates, an iteration of adaptive smoothing yields

$$I^{(t+1)}(\vec{x}) = \frac{\sum_{i=-1}^{+1} \sum_{j=-1}^{+1} I^{(t)}(x_1 + i, x_2 + j) w^{(t)}}{\sum_{i=-1}^{+1} \sum_{j=-1}^{+1} w^{(t)}}, \tag{1}$$

where the convolution mask $w^{(t)}$ is defined as

$$w^{(t)}(x_1, x_2) = \exp\left(-\frac{|d^{(t)}(x_1, x_2)|^2}{2k^2}\right), \tag{2}$$

where k is the variance of the Gaussian mask. In Ref. [21], $d^{(t)}(x_1, x_2)$ is chosen to depend on the magnitude of the gradient computed in a 3×3 window

$$d^{(t)}(x_1, x_2) = \sqrt{G_{x_1}^2 + G_{x_2}^2}, \tag{3}$$

where

$$(G_{x_1}, G_{x_2}) = \left(\frac{\partial I^{(t)}(x_1, x_2)}{\partial x_1}, \frac{\partial I^{(t)}(x_1, x_2)}{\partial x_2}\right), \tag{4}$$

noting the similarity (see also Refs. [3,13,18] for further analogies) of the convolution mask with the diffusion coefficient in anisotropic diffusion [19,26], or more specifically, the total variation in Rudin–Osher–Fatemi's original work [20] that demonstrated how edge-preserving smoothing can be achieved from energy minimization.

2.1. Smoothing on 1D 3-neighborhood

It was suggested in Ref. [21] that Eq. (1) is an implementation of anisotropic diffusion. Briefly sketched, lets consider the case of a one-dimensional signal $I^t(x)$ and reformulate the averaging process as follows

$$I^{t+1}(x) = c_1 I^t(x-1) + c_2 I^t(x) + c_3 I^t(x+1), \tag{5}$$

with

$$c_1 + c_2 + c_3 = 1. \tag{6}$$

Therefore, it is possible to write the above iteration scheme as:

$$I^{t+1}(x) - I^t(x) = c_1(I^t(x-1) - I^t(x)) + c_3(I^t(x+1) - I^t(x)). \tag{7}$$

Taking $c_1 = c_3 = c$, this reduces to

$$I^{t+1}(x) - I^t(x) = c(I^t(x-1) - 2I^t(x) + I^t(x+1)), \tag{8}$$

which is a discrete approximation of the linear diffusion equation:

$$\frac{\partial I}{\partial t} = c\nabla^2 I. \tag{9}$$

2.2. Smoothing on 1D 5-neighborhood

The averaging process can be extended to include second-neighbors

$$I^{t+1}(x) = c_1 I^t(x-2) + c_2 I^t(x-1) + c_3 I^t(x) + c_4 I^t(x+1) + c_5 I^t(x+2), \tag{10}$$

with

$$c_1 + c_2 + c_3 + c_4 + c_5 = 1. \tag{11}$$

Taking $c_1 = c_5 = w_2$, $c_2 = c_4 = w_1$, and $c_3 = 1 - 2w_2 - 2w_1$ this reduces to

$$I^{t+1}(x) = w_2(I^t(x-2) + I^t(x+2)) + (1 - 2w_2 - 2w_1)I^t(x) + w_1(I^t(x-1) + I^t(x+1)), \tag{12}$$

rearrangement of terms leads to

$$I^{t+1}(x) - I^t(x) = w_2(I^t(x-2) - 2I^t(x) + I^t(x+2) + w_1(I^t(x-1) - 2I^t(x) + I^t(x+1)), \tag{13}$$

which is a discrete approximation of the linear diffusion equation

$$\frac{\partial I}{\partial t} = w_1\nabla_1^2 I + w_2\nabla_2^2 I, \tag{14}$$

where ∇_1 denotes ∇ over a grid containing only the nearest-neighbors, and ∇_2 denotes ∇ over a grid containing only the second-neighbors. Typically $w_1 > w_2$ since nearest-neighbors have more influence than second-neighbors.

2.3. Adaptive smoothing on 1D 5-neighborhood

When the weights are space-dependent, one should write the weighted averaging scheme (see Ref. [2] for adaptive smoothing on 1D 3-neighborhood that results in consistency

to the diffusion equation) as follows

$$I^{t+1}(x) = \left(\frac{c^t(x-2)+c^t(x)}{2}\right)I^t(x-2) + \left(\frac{c^t(x-1)+c^t(x)}{2}\right)I^t(x-1) + c^t(x)I^t(x) + \left(\frac{c^t(x+1)+c^t(x)}{2}\right)I^t(x+1) + \left(\frac{c^t(x+2)+c^t(x)}{2}\right)I^t(x+2), \quad (15)$$

with

$$\frac{c^t(x-2)+c^t(x)}{2} + \frac{c^t(x-1)+c^t(x)}{2} + c^t(x) + \frac{c^t(x+1)+c^t(x)}{2} + \frac{c^t(x+2)+c^t(x)}{2} = 1. \quad (16)$$

Plugging Eq. (16) into Eq. (15) and rearranging leads to

$$I^{t+1}(x) - I^t(x) = \frac{c^t(x-2)+c^t(x)}{2}[I^t(x-2) - I^t(x)] + \frac{c^t(x-1)+c^t(x)}{2}[I^t(x-1) - I^t(x)] + \frac{c^t(x+1)+c^t(x)}{2}[I^t(x+1) - I^t(x)] + \frac{c^t(x+2)+c^t(x)}{2}[I^t(x+2) - I^t(x)], \quad (17)$$

or

$$I^{t+1}(x) - I^t(x) = \frac{c^t(x+2)+c^t(x)}{2}[I^t(x+2) - I^t(x)] - \frac{c^t(x-2)+c^t(x)}{2}[I^t(x) - I^t(x-2)] + \frac{c^t(x+1)+c^t(x)}{2}[I^t(x+1) - I^t(x)] - \frac{c^t(x-1)+c^t(x)}{2}[I^t(x) - I^t(x-1)], \quad (18)$$

which is a consistent implementation of the nonlinear diffusion equation

$$\frac{\partial I}{\partial t} = \nabla_1(w_1(x_1,x_2)\nabla_1 I) + \nabla_2(w_2(x_1,x_2)\nabla_2 I), \quad (19)$$

where we have used w instead of c, since the variable w was adopted in Eq. (14) instead of c in Eq. (9). Thus, the weights $w_1(x_1,x_2), w_2(x_1,x_2)$ are the nonlinear diffusion coefficients in the nearest-neighbors grid or second-neighbor grid, respectively, typically taken as

$$w_{1,2}(x_1,x_2) = g(||\nabla_{1,2} I(x_1,x_2)||), \quad (20)$$

where $||\nabla_{1,2} I||$ is the gradient magnitude on either the nearest-neighbors grid or the second-neighbors grid, respectively, and $g(||\nabla_{1,2} I||)$ is an 'edge-stopping' function. This function is chosen to satisfy $g(x) \to 0$ when $x \to \infty$ so that the diffusion is stopped across edges. Thus, a fundamental link between the nonlinear diffusion equation and edge-preserving smoothing filters is noticed (in addition to the interesting relationships between nonlinear PDEs and morphological filters that have been explored in Refs. [5,14,17,25], for example). This link will be extended in Section 3 to $(2S+1 \times 2S+1)$-neighborhood, and will lead to constructing the bilateral filter as a basic mechanism for the mean shift procedure.

3. Generalized adaptive smoothing and nonlinear diffusion on extended neighborhood

Adaptive smoothing was introduced in Ref. [21] as a local process applying a 3×3 window at each iteration, as defined in Eq. (1). However, it is natural to extend this definition to an arbitrary, $(2S+1 \times 2S+1)$ window

$$\vec{I}^{(t+1)}(\vec{x}) = \frac{\sum_{i=-S}^{+S}\sum_{j=-S}^{+S} \vec{I}^{(t)}(x_1+i, x_2+j)w^{(t)}}{\sum_{i=-S}^{+S}\sum_{j=-S}^{+S} w^{(t)}}, \quad (21)$$

where $\vec{I}$ is a three-element vector that describes color images. In the rest of this section, the extended nonlinear diffusion is derived using the generalization of adaptive smoothing outlined in Eq. (21).

First, it is instructive to derive the extended nonlinear diffusion in one-dimension. Considering adaptive smoothing on 1D $(2S+1)$-neighborhood, it is possible to generalize Eq. (15) by using vector notation to

$$I^{t+1}(x) = \left(\frac{\vec{c}_{(-)} + c(x)\hat{1}}{2}\right)\cdot\vec{I}^t_{(-)} + c(x)I^t(x) + \left(\frac{\vec{c}_{(+)} + c(x)\hat{1}}{2}\right)\cdot\vec{I}^t_{(+)}, \quad (22)$$

where $\hat{1} = [1,1,\ldots,1]$ is the unity vector, '$\cdot$' denotes the dot product, $\vec{c}_{(-)} = [c(x-S), c(x-S+1), \ldots, c(x-1)]$, $\vec{c}_{(+)} = [c(x+1), c(x+2), \ldots, c(x+S)]$, $\vec{I}_{(-)} = [I(x-S), I(x-S+1), \ldots, I(x-1)]$, $\vec{I}_{(+)} = [I(x+1), I(x+2), \ldots, I(x+S)]$, $c(x)$ and $I(x)$ are scalars, $I(x)$ being the gray-level intensity at the point of interest x. In color images, I becomes a three-element vector, as in Eq. (21), but this extension is avoided in Eq. (22) for the purpose of clarity and is deferred until the final expression for the two-dimensional case. By analogy to Eq. (16), or adaptive

smoothing in 1D 3-neighborhood outlined in Ref. [2], normalization of the weights can be written in vector notation as

$$\left(\frac{\vec{c}_{(-)}+c(x)\hat{1}}{2}\right)\cdot\hat{1}+c(x)+\left(\frac{\vec{c}_{(+)}+c(x)\hat{1}}{2}\right)\cdot\hat{1}=1, \quad (23)$$

and by analogy to Eqs. (18) or (13) of Ref. [2], we obtain

$$\begin{aligned} I^{t+1}(x)-I^t(x) &= \left(\frac{\vec{c}_{(+)}+c(x)\hat{1}}{2}\right)\cdot[\vec{I}^t_{(+)}-I^t(x)\hat{1}] \\ &\quad -\left(\frac{\vec{c}_{(-)}+c(x)\hat{1}}{2}\right)\cdot[I^t(x)\hat{1}-\vec{I}^t_{(-)}], \end{aligned} \quad (24)$$

which is an implementation of the nonlinear diffusion equation

$$\frac{\partial I(x)}{\partial t}=\vec{\nabla}\cdot(\vec{w}(x)\vec{\nabla}I(x)), \quad (25)$$

where $\vec{\nabla}=[\nabla_{-S},\nabla_{-S+1},\ldots,\nabla_S]$ is a vector containing gradients taken at different neighboring configurations (i.e. nearest-neighbors, second-neighbors, etc.) and $\vec{w}=[w_{-S},w_{-S+1},\ldots,w_S]$ are the nonlinear diffusion coefficients. It is also possible to write Eq. (25) as

$$\frac{\partial I(x)}{\partial t}=\sum_{-S}^{S}\nabla_j(w_j(x)\nabla_j I(x)), \quad (26)$$

expanding the vector notation used in Eq. (25).

Second, the generalization of adaptive smoothing in two-dimensions written in Eq. (21) leads to the extended nonlinear diffusion in two-dimensions by simple analogy to Eq. (25). Taking matrices instead of vectors for ∇, w, and using a three-element vector $\vec{I}$ instead of a scalar I to represent color images leads to the extended nonlinear diffusion

$$\frac{\partial\vec{I}(x_1,x_2)}{\partial t}=\tilde{\nabla}\cdot(\tilde{w}(x_1,x_2)\tilde{\nabla}\vec{I}(x_1,x_2)), \quad (27)$$

where "·" denotes the scalar product between two matrices, and $\tilde{\nabla}$, $\tilde{w}$ are $(2S+1)\times(2S+1))$ matrices that correspond to different neighbor combinations with respect to the center pixel of interest. The generalized adaptive smoothing (21) is a discrete approximation of the *extended nonlinear diffusion* (27). It is noted that in practice, S need not be taken too large (i.e. $S\leq 3$), otherwise the generalized adaptive smoothing becomes an inaccurate representation of the extended diffusion equation.

4. Bilateral filtering

The idea of combining space and color for computer vision tasks has been explored in several works (e.g. Refs. [6,15,16,22,23,29]) and consequently, related digital filters have been proposed in Refs. [7,24]. In this section, the Kimmel–Malladi–Sochen approach [15,23] of using the geometry of spatial–color space to perform edge-preserving smoothing is used to systematically choose the weights of the extended nonlinear diffusion that yields bilateral filtering. Properties of bilateral filtering can be found in Refs. [1,2,11,12,22,24]. In Ref. [22], it was shown that the bilateral filter is closely related to the short time kernel of the Beltrami [23].

Bilateral filtering was introduced [24] as a nonlinear filter which combines domain and range filtering. Given an input image $\vec{I}(\vec{x})$, using a continuous representation notation as in Ref. [24], the output image $\vec{h}(\vec{x})$ is obtained by

$$\vec{h}(\vec{x})=\frac{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\vec{I}(\vec{\xi})c(\vec{\xi},\vec{x})s(\vec{I}(\vec{\xi}),\vec{I}(\vec{x}))\mathrm{d}\vec{\xi}}{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}c(\vec{\xi},\vec{x})s(\vec{I}(\vec{\xi}),\vec{I}(\vec{x}))\mathrm{d}\vec{\xi}}, \quad (28)$$

where $\vec{x}=(x_1,x_2)$, $\vec{\xi}=(\xi_1,\xi_2)$ are space variables and $\vec{I}=(I_R,I_G,I_B)$ is the intensity. The convolution mask is the product of the functions c and s, which represent 'closeness' (in the domain) and 'similarity' (in the range), respectively.

It was demonstrated in Ref. [2] that a discrete version of Gaussian bilateral filtering can be written as follows

$$\vec{I}^{(t+1)}(\vec{x})=\frac{\sum_{i=-S}^{+S}\sum_{j=-S}^{+S}\vec{I}^{(t)}(x_1+i,x_2+j)w^{(t)}}{\sum_{i=-S}^{+S}\sum_{j=-S}^{+S}w^{(t)}}, \quad (29)$$

with the weights given by

$$w^{(t)}(\vec{x},\vec{\xi})=\exp\left(\frac{-(\vec{\xi}-\vec{x})^2}{2\sigma_D^2}\right)\exp\left(\frac{-(I(\vec{\xi})-I(\vec{x}))^2}{2\sigma_R^2}\right), \quad (30)$$

where S is the window size of the filter. Since Eq. (29) and the generalized adaptive smoothing Eq. (21) are equivalent, what remains to be shown is an explanation for the origin of the weights given in Eq. (30).

In color images, it was demonstrated in Refs. [15,23] that the image can be represented as a $2D$ surface embedded in the $5D$ spatial–color space and denoising can be achieved by using the Beltrami flow. For representing the geometry of the $5D$ (x,y,R,G,B) space, it is simple and logical to define the local measure as

$$ds^2=dx^2+dy^2+\beta^2(dR^2+dG^2+dB^2), \quad (31)$$

which is the geometric arclength in the hybrid spatial–color space. Thus, the distance measure in Eq. (2) is given by:

$$|d^{(t)}(x_1,x_2)|^2=\Delta x_1^2+\Delta x_2^2+\beta^2(\Delta R^2+\Delta G^2+\Delta B^2). \quad (32)$$

In Refs. [1,2] a rigorous analysis was worked out, by defining the *generalized intensity* in the $5D$ spatio-colour space, for what can be intuitively conjectured; namely,

plugging Eq. (32) into Eq. (2) yields

$$w^{(t)}(\vec{x},\vec{\xi}) = \exp\left(\frac{-(\vec{\xi}-\vec{x})^2}{2\sigma_D^2}\right)\exp\left(\frac{-(I(\vec{\xi})-I(\vec{x}))^2}{2\sigma_R^2}\right), \quad (33)$$

where $\beta = \sigma_D/\sigma_R$, $\vec{x}$ is the location of the pixel of interest, $\vec{\xi}$ is the location of a pixel in its vicinity inside the window, and

$$\left(I(\vec{x}) - I(\vec{\xi})\right)^2 = (\Delta R)^2 + (\Delta G)^2 + (\Delta B)^2, \quad (34)$$

noting that in Eq. (31), the RGB color space was chosen in defining the geometric arclength for illustrative purposes, but different color spaces of interest such as the CIE Luv or CIE Lab can be chosen for I, depending on the application.

The extended adaptive smoothing or equivalently, the extended nonlinear diffusion (27), along with the choice of weights according to Eq. (33) yields precisely the Gaussian bilateral filter (29), (30) proposed in Ref. [24]. Thus, bilateral filtering is obtained from the extended nonlinear diffusion, with the weights (i.e. the nonlinear diffusivities) chosen according to the geometric arclength in the $5D$ spatio-color space defined by Eq. (31).

5. Mean shift-based filtering

This section introduces the mean shift procedure as an iterative algorithm for local mode detection in the joint spatial–range domain. The mean shift-based filtering is defined in the sequel followed by a discussion on the relation to bilateral filtering.

Let us denote by $\mathbf{x}_i$, $i = 1,\ldots,n$ a set of n data points in the d-dimensional space R^d. The multivariate kernel density estimator with normal kernel and a symmetric positive definite $d \times d$ bandwidth matrix $\mathbf{H}$, computed at the point $\mathbf{x}$ is given by

$$\hat{f}(\mathbf{x}) = \frac{1}{n|2\pi\mathbf{H}|^{1/2}}\sum_{i=1}^{n}\exp\left(-\frac{1}{2}d^2(\mathbf{x},\mathbf{x}_i,\mathbf{H})\right), \quad (35)$$

where

$$d^2(\mathbf{x},\mathbf{x}_i,\ \mathbf{H}) \equiv (\mathbf{x}-\mathbf{x}_i)^\top\mathbf{H}^{-1}(\mathbf{x}-\mathbf{x}_i), \quad (36)$$

is the Mahalanobis distance from $\mathbf{x}$ to $\mathbf{x}_i$. By computing the gradient of $\hat{f}(\mathbf{x})$

$$\nabla\hat{f}(\mathbf{x}) = \frac{\mathbf{H}^{-1}}{n|2\pi\mathbf{H}|^{1/2}}\sum_{i=1}^{n}(\mathbf{x}_i-\mathbf{x})\exp\left(-\frac{1}{2}d^2(\mathbf{x},\mathbf{x}_i,\mathbf{H})\right), \quad (37)$$

after some algebra we have

$$\mathbf{m}(\mathbf{x}) = \mathbf{H}\frac{\nabla\hat{f}(\mathbf{x})}{\hat{f}(\mathbf{x})}, \quad (38)$$

where

$$\mathbf{m}(\mathbf{x}) \equiv \frac{\sum_{i=1}^{n}\mathbf{x}_i\exp\left(-\frac{1}{2}d^2(\mathbf{x},\mathbf{x}_i,\mathbf{H})\right)}{\sum_{i=1}^{n}\exp\left(-\frac{1}{2}d^2(\mathbf{x},\mathbf{x}_i,\mathbf{H})\right)} - \mathbf{x}, \quad (39)$$

is the mean shift vector. Observe that $\mathbf{m}(\mathbf{x})$ is an estimator of the normalized gradient of the underlying density. The repetitive computation of Eq. (39) followed by the translation of the kernel according to the mean shift vector defines a procedure which leads to a local mode of the density [8,9].

Assume now that the data points $\mathbf{x}_i$, are the *generalized pixels* of the input image. This means the vector components of $\mathbf{x}$ contain both the spatial lattice information $\mathbf{z} = (z_1, z_2)^\top$ and range information $\mathbf{c}$, i.e.

$$\mathbf{x}_i = (\mathbf{z}_i^\top, \mathbf{c}_i^\top)^\top, \quad (40)$$

with $i = 1,\ldots,n$. The dimension of vector $\mathbf{c}$ is $r = 1$ when only the intensity values are considered, $r = 3$ for color images, or $r > 3$ in the multispectral case. Although a more complex form of the bandwidth can be useful in certain applications, we will assume henceforth that the bandwidth matrix $\mathbf{H}$ is diagonal having the diagonal terms equal to σ_D^2 for the spatial part and σ_R^2 for the range part. Using these notations, the mean shift vector (39) can be expressed as

$$\mathbf{m}(\mathbf{x}) = \frac{\sum_{i=1}^{n}\mathbf{x}_i\exp\left(-\frac{||\mathbf{z}-\mathbf{z}_i||^2}{2\sigma_D^2}\right)\exp\left(-\frac{||\mathbf{c}-\mathbf{c}_i||^2}{2\sigma_R^2}\right)}{\sum_{i=1}^{n}\exp\left(-\frac{||\mathbf{z}-\mathbf{z}_i||^2}{2\sigma_D^2}\right)\exp\left(-\frac{||\mathbf{c}-\mathbf{c}_i||^2}{2\sigma_R^2}\right)} - \mathrm{x}. \quad (41)$$

Denote by $\mathbf{x}_{j,\mathrm{conv}} = (\mathbf{z}_{j,\mathrm{conv}}^\top, \mathbf{c}_{j,\mathrm{conv}}^\top)^\top$ the convergence point of the iterative mean shift procedure initialized in $\mathbf{x}_j = (\mathbf{z}_j^\top, \mathbf{c}_j^\top)^\top$. By running the procedure for all $j = 1,\ldots,n$, each data point is associated to a local mode in the joint spatial–range domain. The *mean shift-based filtered* image $\mathbf{y}_j, j = 1,\ldots,n$ is defined by the range information carried by the point of convergence

$$\mathbf{y}_j = (\mathbf{z}_j^\top, \mathbf{c}_{j,\mathrm{conv}}^\top)^\top. \quad (42)$$

The algorithm has been first published in Ref. [8] and achieves a high quality discontinuity-preserving filtering, by identifying local modes in the joint domain. The novelty is that the kernel moves iteratively in both spatial and range domains, in contrast to the methods discussed up to now, which maintain a fixed spatial component.

A particular case of mean shift filtering related to bilateral filtering can be obtained by fixing the spatial component of the vectors during iterations. The algorithm will again search for the local mode, but only by evolving in the range domain. We call this variant, *restricted* mean shift filtering.

Fig. 1. Original squirrel image.

An important feature of mean shift filtering is that the image structure does not change during iterations. The algorithm evolution is driven by the initial image structure. By contrast, both nonlinear diffusion and bilateral filtering change the initial image structure and will converge to a flat image, if run until convergence (although, this can be remedied in nonlinear diffusion by introducing an extra term to the diffusion equation as in Ref. [20]). Neither nonlinear diffusion nor the bilateral filter, when performed repeatedly, seek the local mode of the density. Therefore, it is expected that the meanshift mechanism will achieve better image structure preservation, which is demonstrated in Section 6. We note that the principles of mean shift filtering were recently rediscovered in Ref. [28], where the algorithm is called local mode filtering.

6. Experiments

We compare in this section the performance of a simple nonlinear diffusion (as in Ref. [19]), bilateral filtering, and mean shift filtering. The comparison is performed on a B/W squirrel image (Fig. 1) in order to observe the main features of the underlying mechanisms.

Nonlinear diffusion results with $\Delta t = 1.0$ are presented in Fig. 2 after 10, 20, 50 and 100 iterations. In Figs. 3 and 4 we show bilateral filtering with $\sigma_D = 3.0$, $\sigma_R = 25.0$ and $\sigma_D = 3.0$, $\sigma_R = 35.0$, respectively, after 1, 2, 5 and 10 iterations. Restricted mean shift results (kernel moving in range domain) are shown in Fig. 5 for $\sigma_D = 3.0$ and 5.0, $\sigma_R = 25.0$ and 35.0. The same parameters are used to generate the unrestricted mean shift results (kernel moving in both domains), presented in Fig. 6.

The following observations can be derived:

- Nonlinear diffusion obtains a pleasant result after 50 iterations, although many regions around the squirrel tail are excessively smoothed, while the borders are not very well defined. The amount of excessive smoothing increases at 100 iterations.
- Similar comments are valid for bilateral filtering. After 2 iterations a good compromise between the amount of smoothing and preserved edges is reached. Nevertheless,

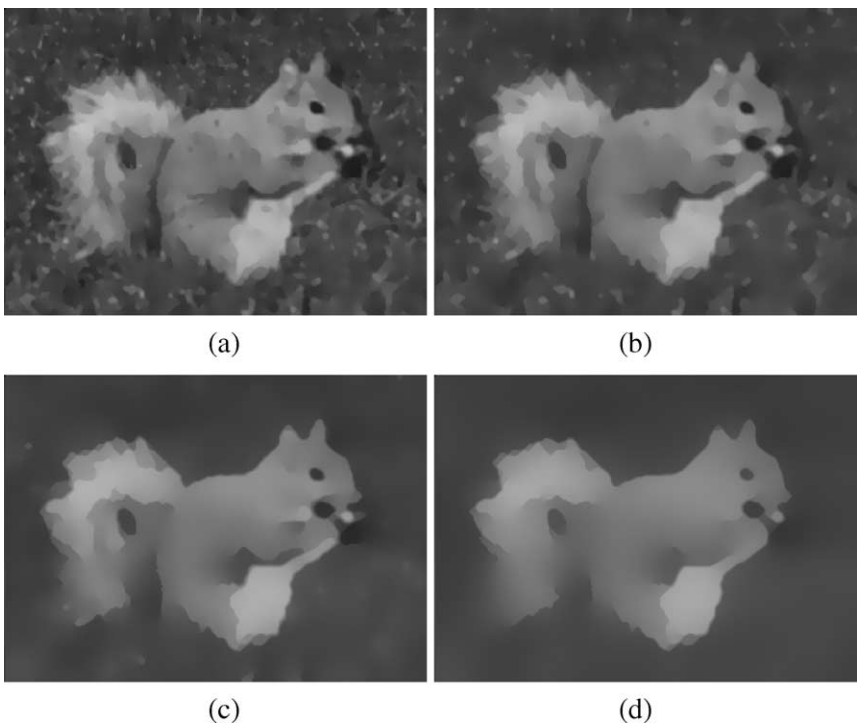

Fig. 2. Nonlinear diffusion with $\Delta t = 1.0$. (a) 10 iterations. (b) 20 iterations. (c) 50 iterations. (d) 100 iterations.

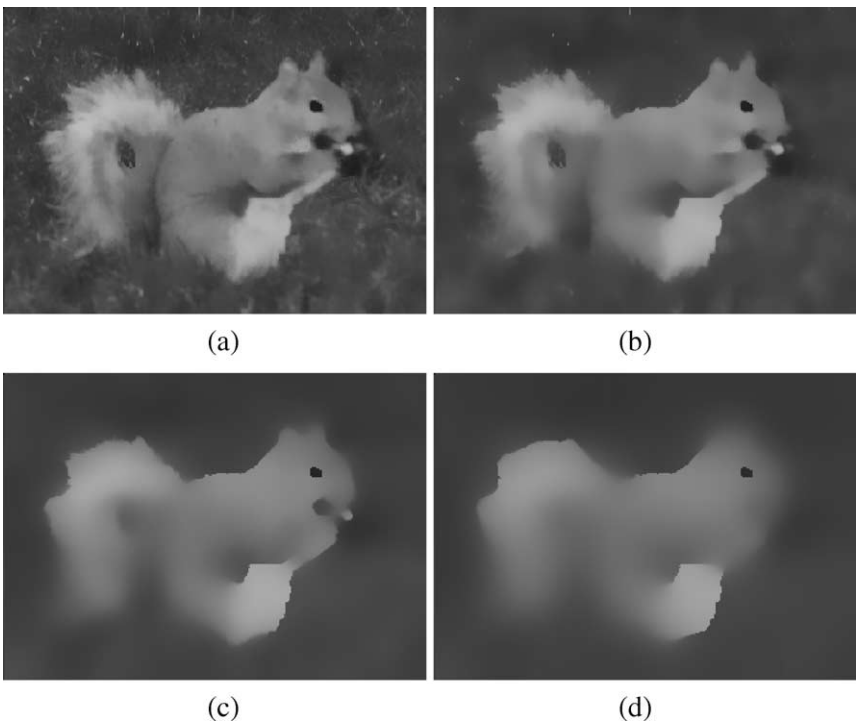

Fig. 3. Bilateral filtering with $\sigma_D = 3.0$ and $\sigma_R = 25.0$. (a) 1 iteration. (b) 2 iterations. (c) 5 iterations. (d) 10 iterations.

excessive smoothing is present around the tail regions. The gradual collapse of the processed data to a flat image is noticeable after five iterations.

- Restricted mean shift filtering bears resemblance to the first two techniques in terms of quality of the preserved edges.
- Unrestricted mean-shift is successful in achieving the sharpest boundaries among all the various approaches examined (see the quality of results in Fig. 6c and d). The reason is that the local structure is better exploited by letting the kernel to simultaneously move in both spatial and range domains.

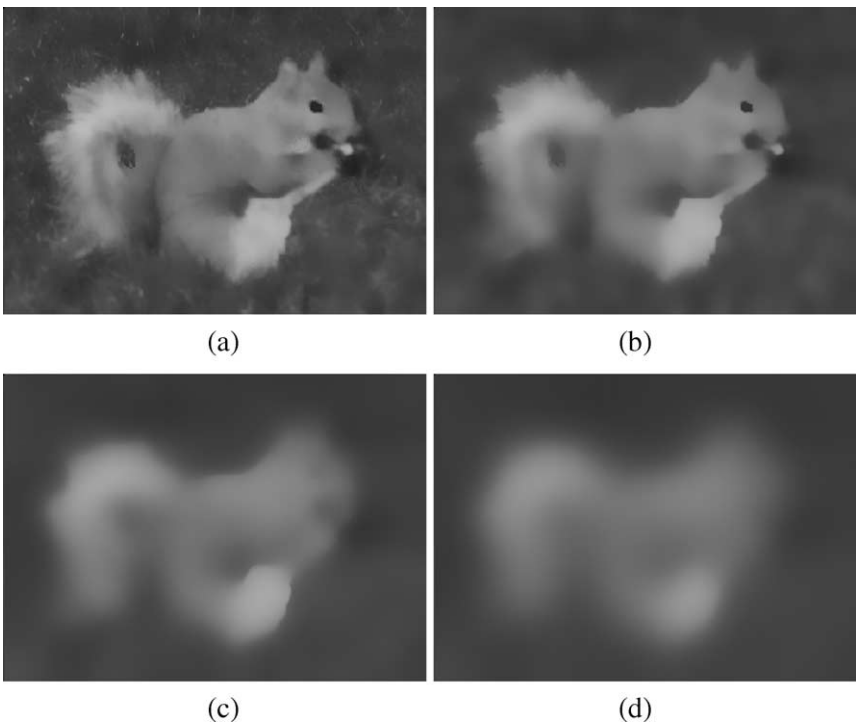

Fig. 4. Bilateral filtering with $\sigma_D = 3.0$ and $\sigma_R = 35.0$. (a) 1 iteration. (b) 2 iterations. (c) 5 iterations. (d) 10 iterations.

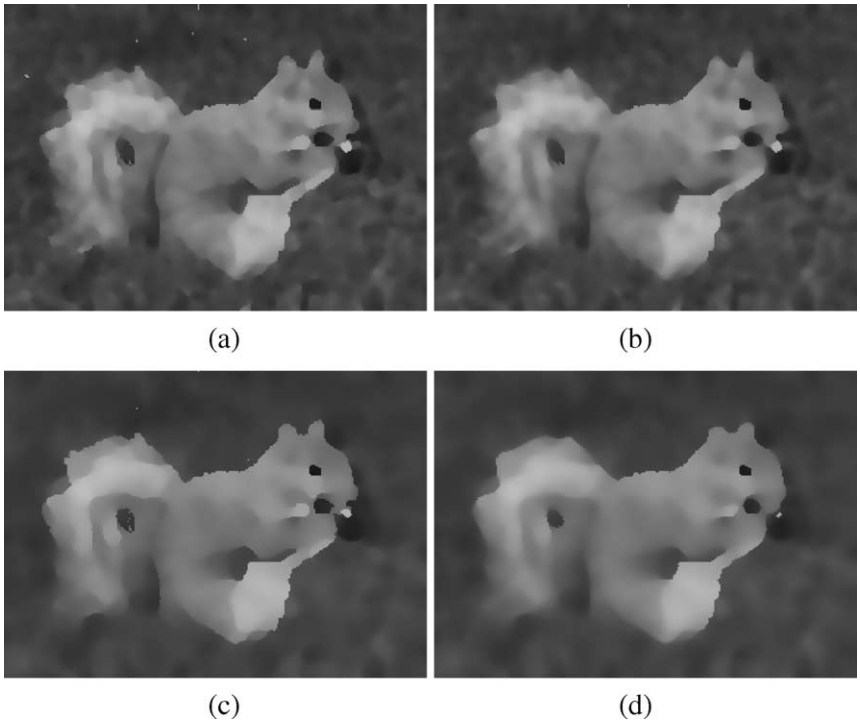

Fig. 5. Restricted mean shift filtering. (a) $\sigma_D = 3.0$, $\sigma_R = 25.0$. (b) $\sigma_D = 3.0$, $\sigma_R = 35.0$. (c) $\sigma_D = 5.0$, $\sigma_R = 25.0$. (d) $\sigma_D = 5.0$, $\sigma_R = 35.0$.

7. Conclusion

A common framework has been formulated for nonlinear diffusion [19,20,26], adaptive smoothing [21], bilateral filtering [24], and the mean shift paradigm [9]. Emphasizing the importance of extended neighborhoods, both nonlinear diffusion and adaptive smoothing can be generalized and unified to a single approach that accomplishes edge-preserving smoothing by using $(2S + 1 \times 2S + 1)$ instead of 3×3 window at each iteration (i.e. an extended nonlinear diffusion). The extended nonlinear diffusion process can then be casted into bilateral filtering by a specific choice of

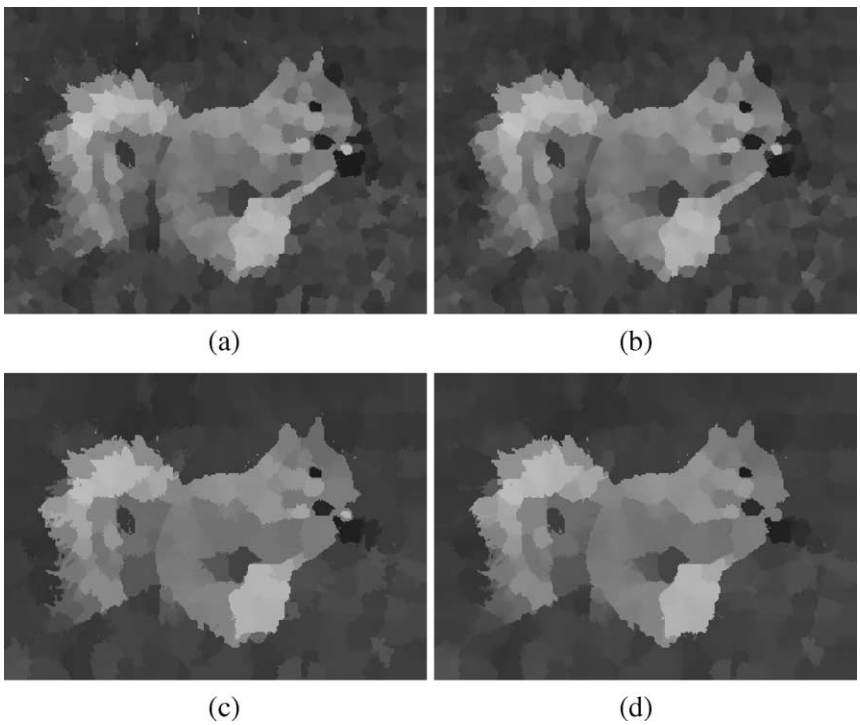

Fig. 6. Mean shift filtering (unrestricted). (a) $\sigma_D = 3.0$, $\sigma_R = 25.0$. (b) $\sigma_D = 3.0$, $\sigma_R = 35.0$. (c) $\sigma_D = 5.0$, $\sigma_R = 25.0$. (d) $\sigma_D = 5.0$, $\sigma_R = 35.0$.

weights, based on geometrical considerations [15,23]. The bilateral mechanism is in turn related to a robust iterative procedure (i.e. the mean shift) which achieves edge-preserving filtering by searching for local modes in the joint spatial–range domain. We have thus established a noteworthy link between the nonlinear diffusion and the kernel methods from statistics. As a result, various tools derived with statistical motivations such as bandwidth selection [10] could be interpreted and exploited for parameter selection in the diffusion process.

References

[1] D. Barash, Bilateral filtering and anisotropic diffusion: towards a unified viewpoint, Hewlett-Packard Laboratories Technical Report, HPL-2000-18(R.1), 2000.

[2] D. Barash, A fundamental relationship between bilateral filtering, adaptive smoothing, and the nonlinear diffusion equation, IEEE Transactions on Pattern Analysis and Machine Intelligence 24 (6) (2002) 844.

[3] M.J. Black, G. Sapiro, D. Marimont, D. Heeger, Robust anisotropic diffusion, IEEE Transactions on Image Processing 7 (3) (1998) 421.

[4] T. Boult, R.A. Melter, F. Skorina, I. Stojmenovic, G-neighbors, Proceedings of the SPIE, Vision Geometry II 2060 (1993) 96.

[5] V. Caselles, G. Sapiro, D.H. Chung, Vector median filters. inf–sup operations, and coupled PDE's: theoretical connections, Journal of Mathematical Imaging and Vision 8 (2000) 109.

[6] T.F. Chan, J. Shen, Variational restoration of non-flat image features: models and algorithms, SIAM Journal of Applied Mathematics 61 (4) (2000) 1338.

[7] T.F. Chan, S. Osher, J. Shen, The digital TV filter and nonlinear denoising, IEEE Transactions on Image Processing 10 (2) (2001) 231.

[8] D. Comaniciu, P. Meer, Mean shift analysis and applications, Proceedings of the 1999 IEEE International Conference on Computer Vision, Kerkyra, Greece, 1999, p. 1197.

[9] D. Comaniciu, P. Meer, Mean shift: a robust approach towards feature space, IEEE Transactions on Pattern Analysis and Machine Intelligence 24 (5) (2002) 603.

[10] D. Comaniciu, An algorithm for data-driven bandwidth selection, IEEE Transactions on Pattern Analysis and Machine Intelligence 25 (2) (2003) 281.

[11] F. Durand, J. Dorsey, Fast bilateral filtering for the display of high-dynamic range image, Proceedings of ACM SIGGRAPH 2002, in Computer Graphics Proceedings, San Antonio, TX, 2002.

[12] M. Elad, On the bilateral filter and ways to improve it, IEEE Transactions on Image Processing 11 (10) (2002) 1141.

[13] B. Fischl, E. Schwartz, Adaptive nonlocal filtering: a fast alternative to anisotropic diffusion for image enhancement, IEEE Transactions on Pattern Analysis and Machine Intelligence 21 (1) (1999) 42.

[14] F. Guichard, J.M. Morel, Image Iterative Smoothing and PDE's, online book, September 2000, final version to appear.

[15] R. Kimmel, R. Malladi, N. Sochen, Images as embedding maps and minimal surfaces: movies, color, and volumetric medical images, Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Puerto Rico, 1997.

[16] J.J. Koenderink, A.J. Van Doorn, The structure of locally orderless images, International Journal of Computer Vision 21 (2/3) (1999) 159.

[17] P. Maragos, F. Meyer, Nonlinear PDEs and numerical algorithms for modeling levelings and reconstruction filters, Scale-Space Theories in Computer Vision, Lecture Notes in Computer Science 1682, 1999, p. 363.

[18] M. Nielsen, L. Florack, R. Deriche, Regularization, scale-space and edge detection filters, Journal of Mathematical Imaging and Vision 7 (4) (1997) 291.

[19] P. Perona, J. Malik, Scale-space and edge detection using anisotropic diffusion, IEEE Transactions on Pattern Analysis and Machine Intelligence 12 (7) (1990) 629.

[20] L.I. Rudin, S. Osher, F. Fatemi, Nonlinear total variation based noise removal algorithms, Physica D 60 (1992) 259.

[21] P. Saint-Marc, J.S. Chen, G. Medioni, Adaptive smoothing: a general tool for early vision, IEEE Transactions on Pattern Analysis and Machine Intelligence 13 (6) (1991) 514.

[22] N. Sochen, R. Kimmel, A.M. Bruckstein, Diffusions and confusions in signal and image processing, Journal of Mathematical Imaging and Vision 14 (3) (2001) 195.

[23] N. Sochen, R. Kimmel, R. Malladi, A geometrical framework for low level vision, IEEE Transactions on Image Processing 7 (3) (1998) 310.

[24] C. Tomasi, R. Manduchi, Bilateral filtering for gray and color images, Proceedings of the 1998 IEEE International Conference on Computer Vision, Bombay, India, 1998.

[25] R. van den Boomgaard, A.W.M. Smeulders, The morphological structure of images, the differential equations of morphological scale-space, IEEE Transactions on Pattern Analysis and Machine Intelligence 16 (11) (1994) 1101.

[26] J. Weickert, Anisotropic diffusion in image processing, Tuebner Stuttgart, 1998, p. 3, ISBN 3-519-02606-6.

[27] J. Weickert, A scheme for coherence-enhancing diffusion filtering with optimized rotation invariance, CVGPR Group Technical Report at the Department of Mathematics and Computer Science, University of Mannheim, Germany, TR 4/2000, 2000.

[28] J. van de Weijer, R. van den Boomgaard, Local mode filtering, Proceedings of the 2001 IEEE Conference on Computer Vision and Pattern Recognition, Hawaii, vol. 2, 2001, p. 428.

[29] A. Yezzi, Modified curvature motion for image smoothing and enhancement, IEEE Transactions on Image Processing 7 (3) (1998) 345.