A Fundamental Relationship between Bilateral Filtering, Adaptive Smoothing, and the Nonlinear Diffusion Equation

Danny Barash, *Member*, *IEEE*

Abstract—In this paper, the relationship between bilateral filtering and anisotropic diffusion is examined. The bilateral filtering approach represents a large class of nonlinear digital image filters. We first explore the connection between anisotropic diffusion and adaptive smoothing, and then the connection between adaptive smoothing and bilateral filtering. Previously, adaptive smoothing was considered an inconsistent approximation to the nonlinear diffusion equation. We extend adaptive smoothing to make it consistent, thus enabling a unified viewpoint that relates between nonlinear digital image filters and the nonlinear diffusion equation.

Index Terms—Bilateral filtering, anisotropic diffusion, adaptive smoothing, denoising.

1 Introduction

In a wide variety of applications, it is necessary to smooth an image while preserving its edges. Simple smoothing operations such as low-pass filtering, which does not take into account intensity variations within an image, tend to blur edges. Anisotropic diffusion [6] was proposed as a general approach to accomplish edge-preserving smoothing. This approach has grown to become a well-established tool in early vision.

This paper examines the relationship between bilateral filtering, a recent approach proposed in [9] that represents a large class of nonlinear digital image filters and anisotropic diffusion. Applications of bilateral filtering are varied (see, for example, the mean shift filtering [3] applications, mean shift and bilateral filtering, are closely related).

The paper is divided as follows: Section 2 presents the connection between anisotropic diffusion and adaptive smoothing, modifying the adaptive smoothing that was previously suggested in [7] to make it consistent. The goal is to suggest a viewpoint in which adaptive smoothing serves as the link between bilateral filtering and anisotropic diffusion. In Section 3, adaptive smoothing is further extended, which results in bilateral filtering. The possible unification of bilateral filtering and anisotropic diffusion is then discussed. Sections 4 and 5 take advantage of the resultant link, suggesting an analogy between the geometric interpretation in anisotropic diffusion and bilateral filtering. Section 4 examines the convolution kernel of a bilateral filter from the standpoint that color images are 2D surfaces embedded in 5D $(\mathrm{x}, \mathrm{y}, \mathrm{R}, \mathrm{G}, \mathrm{B})$ space. In Section 5, conclusions are drawn and suggestions are given for future examination of the proposed unified viewpoint.

2 Anisotropic Diffusion and Adaptive Smoothing

We first examine the connection between anisotropic diffusion and adaptive smoothing, which was outlined in [7]. 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:

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

where k is the variance of the Gaussian mask. In [7], $d^{(t)}(x_1, x_2)$ 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 of the convolution mask with the diffusion coefficient in anisotropic diffusion [6], [10].

It was suggested in [7] that (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 follows:

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

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:

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

However, when the weights are space-dependent, one should write the weighted averaging scheme as follows:

$$\begin{aligned} I^{t+1}(x) = {} & \frac{c^t(x-1)I^t(x-1) + c^t(x)I^t(x-1)}{2} \\ & + c^t(x)I^t(x) \\ & + \frac{c^t(x+1)I^t(x+1) + c^t(x)I^t(x+1)}{2} \end{aligned} \tag{10}$$

with

$$\frac{c^t(x-1) + c^t(x)}{2} + c^t(x) + \frac{c^t(x+1) + c^t(x)}{2} = 1. \tag{11}$$

Note that this is different than the original suggestion in [7]. The Appendix explains why [7] leads to an inconsistent approximation of the nonlinear diffusion equation, as noted in [12], while (10) and (11) manage to fix adaptive smoothing in order to make it consistent. Plugging (11) into (10) and rearranging leads to:

$$\begin{aligned} I^{t+1}(x) - I^t(x) = {} & \frac{c^t(x-1) + c^t(x)}{2}[I^t(x-1) \\ & - I^t(x)] + \frac{c^t(x+1) + c^t(x)}{2}[I^t(x+1) - I^t(x)] \end{aligned} \tag{12}$$

• *The author is with the Department of Chemistry and the Courant Institute of Mathematical Sciences, New York University and the Howard Hughes Medical Institute, 31 Washington Place, Main 1021, New York, NY 10003. E-mail: barash@biomath.nyu.edu.*

Manuscript received 24 July 2001; revised 24 Oct. 2001; accepted 26 Oct. 2001.
Recommended for acceptance by P. Meer.
For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number 114594.

or

$$I^{t+1}(x) - I^t(x) = \frac{c^t(x+1) + c^t(x)}{2}[I^t(x+1) - I^t(x)] - \frac{c^t(x-1) + c^t(x)}{2}[I^t(x) - I^t(x-1)], \tag{13}$$

which is an implementation of anisotropic diffusion, proposed by Perona and Malik [6]:

$$\frac{\partial I}{\partial t} = \nabla(c(x_1, x_2)\nabla I), \tag{14}$$

where $c(x_1, x_2)$ is the nonlinear diffusion coefficient, typically taken as:

$$c(x_1, x_2) = g(\|\nabla I(x_1, x_2)\|), \tag{15}$$

where $\|\nabla I\|$ is the gradient magnitude and $g(\|\nabla I\|)$ 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 link between anisotropic diffusion (14) and adaptive smoothing (1) is established. In the next section, we show the link between adaptive smoothing and bilateral filtering.

3 Bilateral Filtering and Adaptive Smoothing

Bilateral filtering was introduced [9] as a nonlinear filter which combines domain and range filtering. Given an input image $\vec{f}(\vec{x})$, using a continuous representation notation as in [9], the output image $\vec{h}(\vec{x})$ is obtained by:

$$\vec{h}(\vec{x}) = \frac{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \vec{f}(\vec{\xi})c(\vec{\xi},\vec{x})s(\vec{f}(\vec{\xi}),\vec{f}(\vec{x}))d\vec{\xi}}{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} c(\vec{\xi},\vec{x})s(\vec{f}(\vec{\xi}),\vec{f}(\vec{x}))d\vec{\xi}}, \tag{16}$$

where $\vec{x} = (x_1, x_2)$, $\vec{\xi} = (\xi_1, \xi_2)$ are space variables and $\vec{f} = (f_R, f_G, f_B)$ is the intensity. The full vector notation is used in order to avoid confusion in what follows. The convolution mask is the product of the functions c and s, which represent "closeness" (in the domain) and "similarity" (in the range), respectively.

Effectively, we claim that a discrete version of bilateral filtering can be written as follows (using the same notation as in the previous section, only I is now a three-element vector which describes color images):

$$\vec{I}^{(t+1)}(\vec{x}) = \frac{\sum_{i=-S}^{+S}\sum_{j=-S}^{+S} \vec{I}^{(t)}(x_1+i, x_2+j)w^{(t)}}{\sum_{i=-S}^{+S}\sum_{j=-S}^{+S} w^{(t)}} \tag{17}$$

with the weights given by:

$$w^{(t)}(\vec{x},\vec{\xi}) = \exp\left(\frac{-(\vec{\xi}-\vec{x})^2}{2\sigma_D^2}\right)\exp\left(\frac{-(I(\vec{\xi})-I(\vec{x}))^2}{2\sigma_R^2}\right), \tag{18}$$

where S is the window size of the filter, which is a generalization of (1). In order to prove our claim and demonstrate the relation to (1), we use a generalized representation for the intensity $\vec{I}$. In principle, the first element corresponds to the range and the second element corresponds to the domain of the bilateral filter. Defining the generalized intensity as:

$$\widehat{I} \equiv \left\{\frac{\vec{I}(\vec{x})}{\sigma_R}, \frac{\vec{x}}{\sigma_D}\right\}, \tag{19}$$

we now take $d^{(t)}(\vec{x})$ to be the difference between generalized intensities at two points in a given $S \times S$ window, $\left|\widehat{I}(\vec{\xi}) - \widehat{I}(\vec{x})\right|$, the latter being a global extension to (3). In (3), the gradient, being the local difference between two neighboring points in a 3×3 window, was taken as a distance measure. Starting from (2), and setting $k = 1$ since the variances σ_D and σ_R are already included in the generalized intensity, we obtain:

$$\begin{aligned} w^{(t)}(\vec{x}) &= \exp\left(-\frac{1}{2}\left|\widehat{I}(\vec{\xi}) - \widehat{I}(\vec{x})\right|^2\right) \\ &= \exp\left(-\frac{1}{2}\left|\left\{\frac{\vec{I}(\vec{\xi})}{\sigma_R}, \frac{\vec{\xi}}{\sigma_D}\right\} - \left\{\frac{\vec{I}(\vec{x})}{\sigma_R}, \frac{\vec{x}}{\sigma_D}\right\}\right|^2\right) \\ &= \exp\left(-\frac{1}{2}\left|\left\{\frac{\vec{I}(\vec{\xi}) - \vec{I}(\vec{x})}{\sigma_R}, \frac{\vec{\xi}-\vec{x}}{\sigma_D}\right\}\right|^2\right) \\ &= \exp\left(-\frac{1}{2}\left(\frac{(\vec{I}(\vec{\xi}) - \vec{I}(\vec{x}))^2}{\sigma_R^2} + \frac{(\vec{\xi}-\vec{x})^2}{\sigma_D^2}\right)\right) \\ &= \exp\left(\frac{-(\vec{\xi}-\vec{x})^2}{2\sigma_D^2}\right)\exp\left(\frac{-(I(\vec{\xi}) - I(\vec{x}))^2}{2\sigma_R^2}\right). \end{aligned} \tag{20}$$

Because these are the weights used in the bilateral filter, as can be verified in (18), (20) provides a direct link between adaptive smoothing and bilateral filtering. In a general framework of adaptive smoothing, one can take spatial and spectral distance measures along with increasing the window size, abandoning the need to perform several iterations. Taken as such, we get the bilateral filtering implementation of [9] which can be viewed as a generalization of adaptive smoothing.

Fig. 1. Original image prior to edge-preserving smoothing, taken from [9].

4 Geometric Interpretation

In the previous two sections, it was shown that anisotropic diffusion and bilateral filtering can be linked through adaptive smoothing. Specifically, the diffusion coefficient in (14) relates to the convolution mask, in particular to the distance measure which is used in the bilateral filter. Similarly, the relation between anisotropic diffusion and robust statistics was described in [2].

For illustration, Fig. 2 demonstrates two different ways of performing edge-preserving smoothing on the original image in Fig. 1. The result of using nonlinear diffusion filtering and the result of bilateral filtering is similar but not identical since the parameters are different and it was intentionally chosen to use a large window size with the bilateral filter and several iterations with anisotropic diffusion. That is the most natural setup for the two to be used. In some applications, such as the ones described in [10], notions from the nonlinear diffusion equation are borrowed by using several iterations along with a small window size. For such applications, it is clear that a single iteration with a large window size is not adequate and a feedback mechanism is needed. However, in other applications where it is not necessary to closely imitate a diffusion process, for the sake of economical computation, it is possible with one iteration using an extended neighborhood to perform an edge-preserving smoothing that perceives even better. In the specific example that was chosen, one can

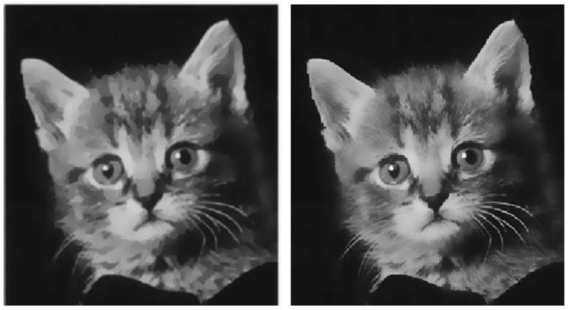

Fig. 2. Edge-preserving smoothing: anisotropic diffusion with 20 time-steps of $\tau = 1.0$ (left) and Gaussian bilateral filtering with a 15×15 window size, $\sigma_D = 5.0$ and $\sigma_R = 30.0$ (right). σ_D and σ_R are bilateral filtering parameters, see [9] for details.

notice the advantage of using a large window size that enables preservation of thin elongated structures such as the whiskers emerging in Fig. 2.

In color images, it was demonstrated in [8] that the image can be represented as 2D surface embedded in the 5D spatial-color space and denoising can be achieved by using the Beltrami flow. Related ideas can be found in [13], [5]. It is possible to borrow this notion outlined in [8] and choose the following spectral distance measure for the bilateral filter:

$$\left|I(\vec{x}) - I(\vec{\xi})\right| = \sqrt{(\Delta R)^2 + (\Delta G)^2 + (\Delta B)^2}, \tag{21}$$

Note that only the spectral distance measure of the range part is given in (21) and can directly be installed in the similarity function s of the bilateral filter as implemented in [9]. Although the RGB color space is chosen here in the range part for illustrative purposes, it is possible to choose other color spaces such as the CIE Luv or CIE Lab depending on the application, as in [9]. The spatial distance measure of the domain part remains the same as with grey-level images. Written that way, one can distinguish between closeness in the domain and similarity in the range, with the advantage of treating the two separately. However, it is also possible to write (21) equivalently by combining the spatial and spectral distance terms. Using the generalized intensity defined in (19), the full distance measure can be written as:

$$\begin{aligned} \left|d^{(t)}(x_1, x_2)\right|^2 &= \left|\sigma_D(\widehat{I(\vec{x})} - \widehat{I(\vec{\xi})})\right|^2 \\ &= (\Delta x_1)^2 + (\Delta x_2)^2 + \beta^2((\Delta R)^2 + (\Delta G)^2 + (\Delta B)^2), \end{aligned} \tag{22}$$

where $\beta = \sigma_D/\sigma_R$. Note that this distance measure can be plugged into the convolution mask of adaptive smoothing (2) as one term with $k = \sigma_D$. It is now possible to take advantage of a geometric interpretation in which color images are $2D$ surfaces embedded in the $5D$ (x, y, R, G, B) space. Equation (22) is then analogous to the local measure:

$$ds^2 = dx^2 + dy^2 + \beta^2(dR^2 + dG^2 + dB^2) \tag{23}$$

which is the geometric arclength in the hybrid spatial-color space discussed in [4], [8].

5 Conclusion

The nature of bilateral filtering resembles that of anisotropic diffusion. It is therefore suggested the two are related and a unified viewpoint can reveal the similarities and differences between the two approaches. Once such an understanding is reached, it is possible to choose the desired ingredients which are common to the two frameworks along with the implementation method. The method can be either applying a nonlinear digital filter or solving a partial-differential equation.

Adaptive smoothing serves as a link between the two approaches, each of which can be viewed as a generalization of the former. In anisotropic diffusion, several iterations of adaptive smoothing are performed. Furthermore, the diffusion coefficient can be generalized to become a "structure tensor" [10] which then leads to phenomena such as edge-enhancing and coherence-enhancing diffusions. In bilateral filtering, the kernel (which plays the same role as the diffusion coefficient) is extended to become globally dependent on intensity, whereas a gradient can only yield local dependency among neighboring pixels. Thus, the window of the filter becomes much bigger in size than the one used in adaptive smoothing and there is no need to perform several iterations. In addition, fast implementations can be used in case the window size is enlarged but remains relatively small. We note that this extension is general on its own right, meaning that a variety of yet unexplored possibilities exist for constructing a kernel with an optimal window size, as well as designing the best closeness and similarity functions for a given application.

The general hybrid spatial-color formulation [4], [8] provide a geometric interpretation with which the bilateral convolution kernel can be viewed as an approximation to the geometric arclength in the 5D hybrid spatial-color space. Ideas that are based on the geometric interpretation, such as coherence-enhancement, can be borrowed from anisotropic diffusion and applied to some degree of approximation in bilateral filtering [1].

Two practical goals seem to come up from comparing between anisotropic diffusion and bilateral filtering. The first is a further trial to reduce the number of iterations needed in anisotropic diffusion (which can be achieved by efficient numerical schemes such as [11], less prone to stability problems). The second is to reduce the window size and investigate other means which aim at minimizing computations associated with bilateral filtering. Both approaches are related to each other and an exchange of new ideas between one another can be rewarding.

Finally, from a conceptual viewpoint, the relationship that was established reveals a basic connection between nonlinear digital image filters and the nonlinear diffusion equation.

Appendix

Let us first examine the consistency problem in the original adaptive smoothing, introduced in [7]. Taking the weights to be space-dependent, the weighted averaging scheme was taken as:

$$\begin{aligned} I^{t+1}(x) = {} & c^t(x-1)I^t(x-1) \\ & + c^t(x)I^t(x) + c^t(x+1)I^t(x+1) \end{aligned} \tag{24}$$

with

$$c^t(x-1) + c^t(x) + c^t(x+1) = 1. \tag{25}$$

This can be rearranged as:

$$\begin{aligned} I^{t+1}(x) - I^t(x) = {} & c^t(x-1)[I^t(x-1) \\ & - I^t(x)] + c^t(x+1)[I^t(x+1) - I^t(x)] \end{aligned} \tag{26}$$

or

$$\begin{aligned} I^{t+1}(x) - I^t(x) = {} & c^t(x+1)[I^t(x+1) \\ & - I^t(x)] - c^t(x-1)[I^t(x) - I^t(x-1)]. \end{aligned} \tag{27}$$

So far, without a loss of generality, we took the discrete time steps and the spatial spacings to be equal to unity (i.e., $\Delta t = 1$ and $\Delta x = 1$). Thus, we can write the previous equation in full notation as:

$$\begin{aligned}\frac{I^{t+1}(x) - I^t(x)}{\Delta t} =& \frac{1}{\Delta x}\left[c^t(x+1)\frac{(I^t(x+1) - I^t(x))}{\Delta x}\right] \\ &- \frac{1}{\Delta x}\left[c^t(x-1)\frac{(I^t(x) - I^t(x-1))}{\Delta x}\right].\end{aligned} \tag{28}$$

It is now possible to expand, using a Taylor series, the terms $c^t(x+1), c^t(x-1)$ on $c^t(x)$ and $I^t(x+1), I^t(x-1)$ on $I^t(x)$, respectively. As noted by [12], after the expansion, an extra term remains that does not disappear as $\Delta t \to 0$ and $\Delta x \to 0$. This results in an inconsistency. To resolve the consistency problem, it is therefore correct to modify adaptive smoothing as appears in equations (10) and (11) instead of equations (24) and (25). The resultant discretization in (13), that approximates the diffusion (14), is the same discretization used in (8) of [11]. That way, one can perform the Taylor expansion as stated above and verify that adaptive smoothing is consistent.

ACKNOWLEDGMENTS

The author would like to thank Ron Kimmel, from the Computer Science Department at the Technion–Israel Institute of Technology, for his assistance in all aspects of this work. He would also like to thank Renato Keshet, Doron Shaked, and all other members of the Image Processing and Computer Vision Group at Hewlett-Packard Laboratories Israel for encouragement and discussions.

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] M.J. Black, G. Sapiro, D. Marimont, and D. Heeger, "Robust Anisotropic Diffusion," *IEEE Trans. Image Processing,* vol. 7, no. 3, pp. 421, 1998.

[3] D. Comaniciu and P. Meer, "Mean Shift Analysis and Applications," *Proc. 1999 IEEE Int'l Conf. Computer Vision,* 1999.

[4] R. Kimmel, R. Malladi, and N. Sochen, "Images as Embedded Maps and Minimal Surfaces: Movies, Color, Texture, and Volumetric Medical Images," *Int'l J. Computer Vision,* vol. 39, no. 2, pp. 111, 2000.

[5] J.J. Koenderink and A.J. Van Doorn, "The Structure of Locally Orderless Images," *Int'l J. Computer Vision,* vol. 21, no. 2/3, pp. 159, 1999.

[6] P. Perona and J. Malik, "Scale-Space and Edge Detection Using Anisotropic Diffusion," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 12, no. 7, pp. 629, 1990.

[7] P. Saint-Marc, J.S. Chen, and G. Medioni, "Adaptive Smoothing: A General Tool for Early Vision," *IEEE Trans. Pattern Analysis and Machine Intelligence* vol. 13, no. 6, pp. 514, June 1991.

[8] N. Sochen, R. Kimmel, and R. Malladi, "A Geometrical Framework for Low Level Vision," *IEEE Trans. Image Processing,* vol. 7, no. 3, p. 310, 1998.

[9] C. Tomasi and R. Manduchi, "Bilateral Filtering for Gray and Color Images," *Proc. 1998 IEEE Int'l Conf. Computer Vision,* 1998.

[10] J. Weickert, *Anisotropic Diffusion in Image Processing.* Tuebner Stuttgart, 1998.

[11] J. Weickert, B.M. ter Haar Romeny, and M. Viergever, "Efficient and Reliable Schemes for Nonlinear Diffusion Filtering," *IEEE Trans. Image Processing,* vol. 7, no. 3, p. 398, 1998.

[12] J. Weickert, Private Communication, 2001.

[13] A. Yezzi, "Modified Curvature Motion for Image Smoothing and Enhancement," *IEEE Trans. Image Processing,* vol. 7, no. 3, pp. 345, 1998.