Bilateral Filtering and Anisotropic Diffusion: Towards a Unified Viewpoint

Danny Barash
HP Laboratories Israel[1]
HPL-2000-18(R.1)
August 1st, 2000*

E-mail: barash@hpli.hpl.hp.com

bilateral filtering, anisotropic diffusion, adaptive smoothing, denoising

Bilateral filtering has recently been proposed as a noniterative alternative to anisotropic diffusion. In both these approaches, images are smoothed while edges are preserved. Unlike anisotropic diffusion, bilateral filtering does not involve the solution of partial differential equations and can be implemented in a single iteration. Despite the difference in implementation, both methods are designed to prevent averaging across edges while smoothing an image. Their similarity suggests they can somehow be linked. Using a generalized representation for the intensity, we show that both can be related to adaptive smoothing. As a consequence, we also show that bilateral filtering can be applied to denoise and coherence-enhance degraded images with approaches similar to anisotropic diffusion.

[1] Technion City, Haifa 32000 Israel

* Internal Accession Date Only
Approved for External Publication

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 [3] 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 relation between bilateral filtering, a recent approach proposed in [6], and anisotropic diffusion. The paper is divided as follows. Section II presents the connection between anisotropic diffusion and adaptive smoothing. The goal is to suggest a viewpoint in which adaptive smoothing serves as the link between bilateral filtering and anisotropic diffusion. In Section III, adaptive smoothing is extended, which results in bilateral filtering. The possible unification of bilateral filtering and anisotropic diffusion is then discussed. Sections IV and V take advantage of the resultant link, borrowing the use of the geometric interpretation to anisotropic diffusion and applying it in bilateral filtering. Section IV examines the convolution kernel of a bilateral filter, from the standpoint that color images are 2D surfaces embedded in 5D (x,y,R,G,B) space. Consequently, Section V describes another idea from anisotropic diffusion, that of coherence-enhancement of color images using a bilateral filter. In Section VI, 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 [4]. 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 [4], $d^{(t)}(x_1, x_2)$ 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 [3], [7].

It was shown [4] that equation (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:

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

Taking $c_1 = c_3$, this reduces to:

$$I^{t+1}(x) \quad - \quad I^t(x) = c_1(I^t(x-1) - 2I^t(x) + I^t(x+1)) \tag{8}$$

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:

$$I^{t+1}(x) \quad = \quad c^t(x-1)I^t(x-1) + c^t(x)I^t(x) + c^t(x+1)I^t(x+1) \tag{10}$$

with

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

This can be rearranged as:

$$I^{t+1}(x) \quad - \quad I^t(x) = c^t(x-1)(I^t(x-1) - I^t(x)) + c^t(x+1)(I^t(x+1) - I^t(x)) \tag{12}$$

or

$$I^{t+1}(x) \quad - \quad I^t(x) = c^t(x+1)(I^t(x+1) - I^t(x)) - c^t(x-1)(I^t(x) - I^t(x-1)) \tag{13}$$

which is an implementation of anisotropic diffusion, proposed by Perona and Malik [3]:

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

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 [6] 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 [6], 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 3-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:

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

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:

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

we now take $d^{(t)}(\vec{x})$ to be the difference between generalized intensities at two points in a given $S \times S$ window, $\left|\widehat{\vec{I}}(\vec{\xi}) - \widehat{\vec{I}}(\vec{x})\right|$, 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:

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

Because these are the weights used in the bilateral filter, as can be verified in (18), equation (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 [6] which can be viewed as a generalization of adaptive smoothing.

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 and 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 [1].

For illustration, Figure 2 demonstrates two different ways of performing edge-preserving smoothing on the original image in Figure 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.

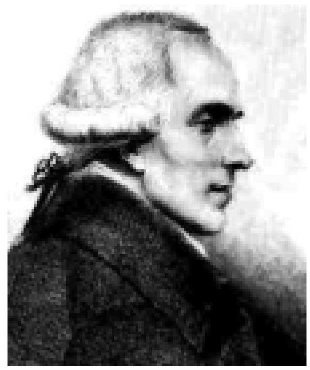

Figure 1: Original image: Laplace.

Figure 2: Edge-preserving smoothing: anisotropic diffusion with 20 time-steps of $\tau = 1.0$ (left) and Gaussian bilateral filtering with a 30×30 window size, $\sigma_D = 5.0$ and $\sigma_R = 30.0$ (right). σ_D and σ_R are bilateral filtering parameters, see [6] for details.

In color images, it was demonstrated in [5] 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. It is possible to borrow this notion outlined in [5] and choose the following spectral distance measure for the bilateral filter:

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

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 [6]. 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:

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

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 [2], [5].

5 Coherence-Enhancement by Bilateral Filtering

In this section, a coherence-enhancement procedure based on the geometric interpretation given in (22) and outlined in [2], [7] is applied using a bilateral filter. The approach, originally proposed by Weickert for anisotropic diffusion, is based on the idea that the amount and direction of diffusion can be controlled by altering the image metric. Our goal is to take a familiar procedure in anisotropic diffusion and apply it in the framework of bilateral filtering in order to achieve control on edge-preservation.

Let us define a difference operator, $\Delta R(x_1, x_2) \equiv |R(x_1, \xi_2) - R(x_1, x_2)| + |R(\xi_1, x_2) - R(x_1, x_2)|$, where the same definition applies to ΔG and ΔB. The coordinates $\vec{x} = (x_1, x_2)$ represent the center of the filter mask, and $\vec{\xi} = (\xi_1, \xi_2)$ is a point within the mask. Furthermore, let $\Delta R_x \equiv |R(x_1, \xi_2) - R(x_1, x_2)| / \Delta x + |R(\xi_1, x_2) - R(x_1, x_2)| / \Delta x$, where Δx stands for either $\Delta x_1 = |x_1 - \xi_1|$ or $\Delta x_2 = |x_2 - \xi_2|$. Using these definitions, one can write the following identity which resembles the chain-rule:

$$\Delta R(x_1, x_2) = \frac{1}{2}(\Delta R_{x_1}\Delta x_1 + \Delta R_{x_2}\Delta x_2) \tag{24}$$

where the same identity applies to ΔG and ΔB. We note that the above definition of the difference operator instead of the straight-forward $\Delta R = |R(\xi_1, \xi_2) - R(x_1, x_2)|$ is meant to give a two-point support to (x_1, x_2) when calculating ΔR_x, which is a better approximation then a single-point support in our attempt to mimic the derivative operator in the context of a geometric framework. It is now possible to write the distance measure of (22) as:

$$\left| d^{(t)}(x_1, x_2) \right|^2 = (\Delta x_1 \quad \Delta x_2) \begin{pmatrix} a & b \\ b & c \end{pmatrix} \begin{pmatrix} \Delta x_1 \\ \Delta x_2 \end{pmatrix} \tag{25}$$

where,

$$\begin{aligned} a &= 1 + \frac{\beta^2(\Delta R_{x_1}^2 + \Delta G_{x_1}^2 + \Delta B_{x_1}^2)}{4} \\ b &= \frac{\beta^2(\Delta R_{x_1}\Delta R_{x_2} + \Delta G_{x_1}\Delta G_{x_2} + \Delta B_{x_1}\Delta B_{x_2})}{4} \\ c &= 1 + \frac{\beta^2(\Delta R_{x_2}^2 + \Delta G_{x_2}^2 + \Delta B_{x_2}^2)}{4} \end{aligned} \tag{26}$$

which can be verified by starting from (25), using the definitions for $\Delta R_{x_1}, \Delta R_{x_2}$ and identity in (24) to reach back (22). We obtain in (25) a symmetric positive definite matrix M analogous to a 'structure tensor' [7]. Diagonalizing the matrix M, we get:

$$M = U\Lambda U^T \tag{27}$$

where $U = (u_1|u_2)$ is a Hermitian matrix with the eigenvectors u_1, u_2 in its columns. The diagonal matrix Λ consists of the eigenvalues λ_1, λ_2 $(\lambda_1 > \lambda_2)$ in its diagonal. For a $2x2$ system which is symmetric (such as the matrix M, see (25)) the corresponding eigenvalues are given by:

$$\lambda_{1,2} = \frac{1}{2}(a + c \pm \sqrt{(a-c)^2 + 4b^2}) \tag{28}$$

and the orthonormal eigenvectors, excluding the trivial case where $b = 0$, can be obtained from:

$$u_1 \parallel \begin{pmatrix} 2b \\ c - a + \sqrt{(a-c)^2 + 4b^2} \end{pmatrix} \tag{29}$$

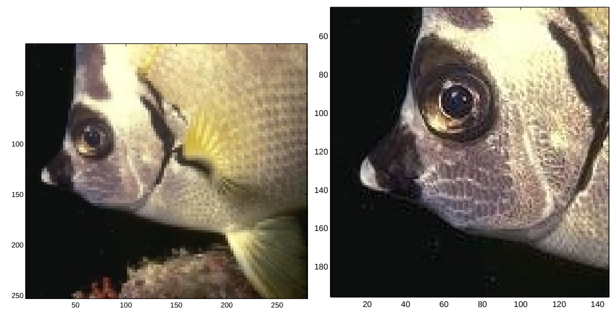

Figure 3: Original Image (this is a color image)

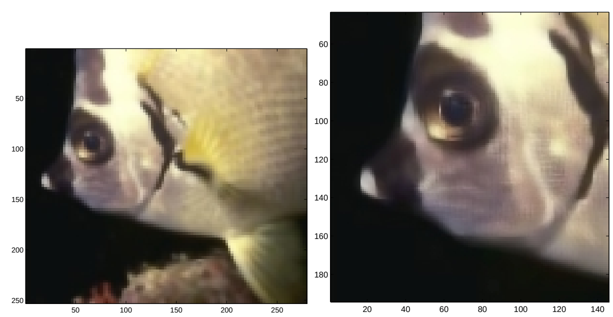

Figure 4: Bilateral Filtering (this is a color image)

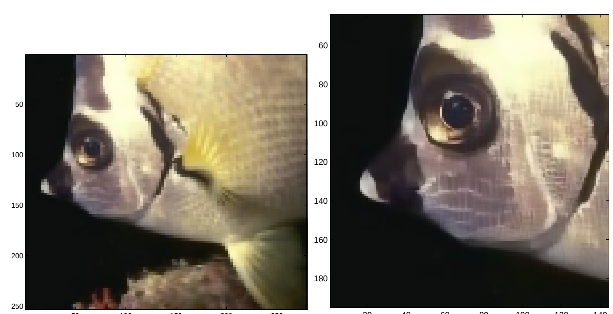

Figure 5: Bilateral Coherence Enhancement (color image)

In our example, we change the eigenvalues of M without altering its eigenvectors, in the same manner used in [2] for the purpose of coherence-enhancement:

$$\Lambda = \begin{pmatrix} \lambda_1 \alpha^{-1} & 0 \\ 0 & \lambda_2 \alpha \end{pmatrix} \tag{30}$$

where $\alpha \ll 1$ is a positive scalar. Unlike the procedure for coherence- enhancement in anisotropic diffusion, diffusing Λ before changing its eigenvalues is not necessary in bilateral filtering.

In the example, we start from an original image of a fish plotted in Figure 3. The image in Figure 4 is then obtained after applying bilateral filtering with $\sigma_D = 10.0$ and $\sigma_R = 100.0$, using the distance measure as in (21). Note that the scales of the fish got smoothed as a consequence of bilateral filtering. In Figure 5, the same bilateral filtering is applied and in addition α is set to 0.1. One can verify by examining the color images that smoothing can be controlled by the variable α. We note that in addition to coherence-enhancement demonstrated here, image enhancement using bilateral filtering can be thought of. Although the framework of bilateral filtering does not allow color image enhancement by inverse diffusion across an edge as mentioned in [2], since the bilateral filter coefficients will always remain positive even when the eigenvalues of the 'structure tensor' are made negative, other approaches can be suggested. One possible approach is to design special closeness and similarity functions, tailored to obtain image enhancement. Another approach which can be considered is to multiply the coherence-enhancing bilateral filter by an additional enhancement factor which attempts to capture the geometry of an edge. Image enhancement by bilateral filtering remains an open and challenging area of application.

6 Conclusions

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 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, the diffusion coefficient can be generalized to become a 'structure tensor' [7] 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. 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 [2], [5] 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.

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 [8], less proned to stability problems) while retaining the same accuracy as in bilateral filtering. 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.

7 Acknowledgment

The author would like to thank Renato Keshet, Michael Elad, Doron Shaked, Ron Maurer and Ron Kimmel for extensively reviewing and commenting on this report.

8 References

[1] Black M.J., Sapiro G., Marimont D., Heeger D., "Robust Anisotropic Diffusion," *IEEE Transactions on Image Processing*, Vol. 7, No. 3, p.421, 1998.

[2] R. Kimmel, R. Malladi, N. Sochen, "Images as Embedded Maps and Minimal Surfaces: Movies, Color, Texture, and Volumetric Medical Images," *International Journal of Computer Vision*, in press.

[3] P. Perona and J. Malik, "Scale-Space and Edge Detection Using Anisotropic Diffusion," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 12, No. 7, p.629, 1990.

[4] P. Saint-Marc, J.S. Chen, G. Medioni, "Adaptive Smoothing: A General Tool for Early Vision," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 13, No. 6, p.514, 1991.

[5] N. Sochen, R. Kimmel, R. Malladi, "A Geometrical Framework for Low Level Vision," *IEEE Transactions on Image Processing*, Vol. 7, No. 3, p.310, 1998.

[6] C. Tomasi and R. Manduchi, "Bilateral Filtering for Gray and Color Images," *Proceedings of the 1998 IEEE International Conference on Computer Vision* , Bombay, India, 1998.

[7] J. Weickert, *Anisotropic Diffusion in Image Processing,* Tuebner Stuttgart, 1998. ISBN 3-519-02606-6.

[8] J. Weickert, B.M. ter Haar Romeny, M. Viergever, "Efficient and Reliable Schemes for Nonlinear Diffusion Filtering," *IEEE Transactions on Image Processing*, Vol. 7, No. 3, p.398, 1998.