Higher Order Bilateral Filters and Their Properties

Hiroyuki Takeda[a], Sina Farsiu[a], and Peyman Milanfar[a]

[a]Electrical Engineering Department, University of California,
1156 High St., Santa Cruz, CA, USA

ABSTRACT

Bilateral filtering[1,2] has proven to be a powerful tool for adaptive denoising purposes. Unlike conventional filters, the bilateral filter defines the closeness of two pixels not only based on geometric distance but also based on radiometric (graylevel) distance. In this paper, to further improve the performance and find new applications, we make contact with a classic non-parametric image reconstruction technique called kernel regression,[3] which is based on local Taylor expansions of the regression function. We extend and generalize the kernel regression method and show that bilateral filtering is a special case of this new class of adaptive image reconstruction techniques, considering a specific choice for weighting kernels and zeroth order Taylor approximation. We show improvements over the classic bilateral filtering can be achieved by using higher order local approximations of the signal.

1. INTRODUCTION

The bilateral filter was first proposed in Ref. 1 as a very effective one pass adaptive filter for denoising purposes while keeping the edges relatively sharp. Since its advent, the idea of bilateral filtering has been modified and improved, and its relation to some of the most popular image enhancement and reconstruction algorithms has been established.To take a notable example, Elad[2] proved that such filter is a single Jacobi iteration of a weighted least squares minimization, and suggested using more iterations to enhance the smoothing effect of estimation. In related works, Barash et. al.[4,5] established the analogy between the geometric interpretation in anisotropic diffusion, mean shift algorithm, and bilateral filtering.

To reduce computational complexity and improve compression efficiency, a modified implementation of bilateral filter is recently proposed in Ref. 6. In another venue, application of bilateral filtering for color image enhancement has been studied in Refs. 7,8. Moreover, a few papers have reported on a specific class of Maximum a Posteriori (MAP) image reconstruction techniques[9,10] that benefit from coupling bilateral filtering with Total Variation[11] defining very efficient adaptive regularization terms.

The data-adaptive kernel regression framework, proposed in Ref. 12 based on the classic kernel regression framework,[3] has wide ranging applications for image and video processing; for example, image denoising, and image interpolation/reconstruction from regularly and irregularly sampled data (i.e. image upscaling, image fusion, and super-resolution). In Ref. 12, we found that the bilateral filter[1,2] is a special case of data-adaptive kernel regression which is called *bilateral kernel regression* (BKR) or *higher order bilateral filter*. The purpose of this paper is to review BKR and examine its performance and behavior numerically and visually in filtered images.

Further author information:
H.T. : E-mail: htakeda@soe.ucsc.edu
S.F. : E-mail: farsiu@soe.ucsc.edu
P.M. : E-mail: milanfar@soe.ucsc.edu
This work was supported in part by AFOSR Grant F49620-03-1-0387.

2. HIGHER ORDER BILATERAL FILTERS

The kernel regression (KR) framework[3] defines its data model in 2-D as

$$y_i = z(\mathbf{x}_i) + \varepsilon_i, \quad i = 1, \cdots, P, \quad \mathbf{x}_i = [x_{1i}, x_{2i}]^T, \tag{1}$$

where y_i is a noisy sample at $\mathbf{x}_i$, $z(\cdot)$ is the (hitherto unspecified) *regression function* to be estimated, ε_i is an i.i.d zero mean noise, and P is the total number of samples in a neighborhood (window) of interest. As such, the kernel regression framework provides a rich mechanism for computing point-wise estimates of the regression function with minimal assumptions about global signal or noise models.

While the specific form of $z(\cdot)$ may remain unspecified, we can rely on a generic local expansion of the function about a sampling point $\mathbf{x}_i$. Specifically, if $\mathbf{x}$ is near the sample at $\mathbf{x}_i$, we have the $(N+1)$-term Taylor series*

$$\begin{aligned} z(\mathbf{x}_i) &\approx z(\mathbf{x}) + \{\boldsymbol{\nabla} z(\mathbf{x})\}^T(\mathbf{x}_i - \mathbf{x}) + \frac{1}{2}(\mathbf{x}_i - \mathbf{x})^T\{\boldsymbol{\mathcal{H}} z(\mathbf{x})\}(\mathbf{x}_i - \mathbf{x}) + \cdots \\ &= \beta_0 + \boldsymbol{\beta}_1^T(\mathbf{x}_i - \mathbf{x}) + \boldsymbol{\beta}_2^T \operatorname{vech}\left\{(\mathbf{x}_i - \mathbf{x})(\mathbf{x}_i - \mathbf{x})^T\right\} + \cdots, \end{aligned} \tag{2}$$

where $\boldsymbol{\nabla}$ and $\boldsymbol{\mathcal{H}}$ are the gradient (2×1) and Hessian (2×2) operators, respectively, and $\operatorname{vech}(\cdot)$ is the half-vectorization operator which lexicographically orders the lower triangular portion of a symmetric matrix. Furthermore, β_0 is $z(\mathbf{x})$, which is the pixel value of interest, and the vectors $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are

$$\boldsymbol{\beta}_1 = \begin{bmatrix} \dfrac{\partial z(\mathbf{x})}{\partial x_1} & \dfrac{\partial z(\mathbf{x})}{\partial x_2} \end{bmatrix}^T, \tag{3}$$

$$\boldsymbol{\beta}_2 = \begin{bmatrix} \dfrac{\partial^2 z(\mathbf{x})}{2\partial x_1^2} & \dfrac{\partial^2 z(\mathbf{x})}{\partial x_1 \partial x_2} & \dfrac{\partial^2 z(\mathbf{x})}{2\partial x_2^2} \end{bmatrix}^T. \tag{4}$$

Since this approach is based on *local* approximations, a logical step to take is to estimate the parameters $\{\boldsymbol{\beta}_n\}_{n=0}^N$ from all the samples $\{y_i\}_{i=1}^P$ while giving the nearby samples higher weights than samples farther away. A formulation of the fitting problem capturing this idea is to solve the following optimization problem,

$$\min_{\{\boldsymbol{\beta}_n\}_{n=0}^N} \sum_{i=1}^{P} \left[y_i - \beta_0 - \boldsymbol{\beta}_1^T(\mathbf{x}_i - \mathbf{x}) - \boldsymbol{\beta}_2^T \operatorname{vech}\left\{(\mathbf{x}_i - \mathbf{x})(\mathbf{x}_i - \mathbf{x})^T\right\} - \cdots \right]^2 K_{\mathbf{H}}(\mathbf{x}_i - \mathbf{x}) \tag{5}$$

where N is the regression order, $K_{\mathbf{H}}(\cdot)$ is the kernel function (a radially symmetric function), $\mathbf{H}(= h\mathbf{I})$ is the (2×2) smoothing matrix, and h is the global smoothing parameter. We control the strength of the smoothing effect by h.

For images, the signal of interest often has discontinuities. Although one way to ensure that the estimated signal approximates such discontinuities is to set N large, the larger N we choose, the dramatically higher the computational complexity of the resulting algorithm becomes. Thus, without setting N large, we add an extra parameter to the kernel function, namely radiometric distance. That yields BKR:

$$\min_{\{\boldsymbol{\beta}_n\}_{n=0}^N} \sum_{i=1}^{P} \left[y_i - \beta_0 - \boldsymbol{\beta}_1^T(\mathbf{x}_i - \mathbf{x}) - \boldsymbol{\beta}_2^T \operatorname{vech}\left\{(\mathbf{x}_i - \mathbf{x})(\mathbf{x}_i - \mathbf{x})^T\right\} - \cdots \right]^2 K_{\mathbf{H}_s}(\mathbf{x}_i - \mathbf{x}) K_{h_r}(y_i - y), \tag{6}$$

where $K_{\mathbf{H}_s}(\mathbf{x}_i - \mathbf{x}) K_{h_r}(y_i - y)$ is specifically a product of two kernels: spatial kernel and radiometric kernel. $\mathbf{H}_s$ $(= h_s\mathbf{I})$ is the 2×2 smoothing matrix, h_s is the spatial smoothing parameter, and h_r is the radiometric smoothing parameter. These two smoothing parameters dictate the "footprint" of the bilateral kernel function.

The optimization problem (6) can be expressed in matrix form as a weighted least-squares optimization problem,[13,14]

$$\begin{aligned} \hat{\mathbf{b}} &= \arg\min_{\mathbf{b}} \left\| \mathbf{y} - \mathbf{X}_{\mathbf{x}}\mathbf{b} \right\|_{\mathbf{W}_{\mathbf{x}}}^2 \\ &= \arg\min_{\mathbf{b}} \left(\mathbf{y} - \mathbf{X}_{\mathbf{x}}\mathbf{b}\right)^T \mathbf{W}_{\mathbf{x}} \left(\mathbf{y} - \mathbf{X}_{\mathbf{x}}\mathbf{b}\right), \end{aligned} \tag{7}$$

*Other localized representations are also possible and may be advantageous.

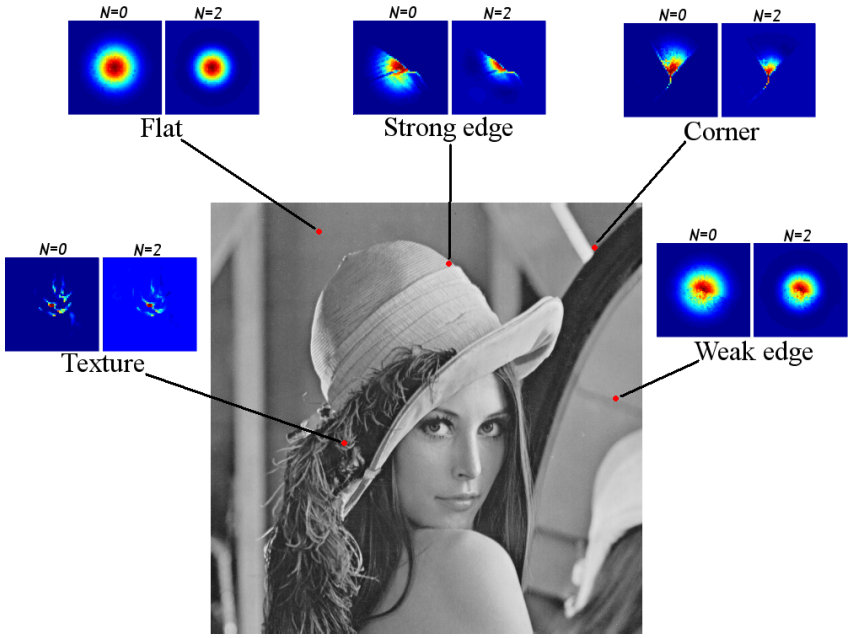


Figure 1. The footprints of the bilateral equivalent weight function $W_i(\mathbf{x}_i - \mathbf{x}, y_i - y, N)$ at a variety of image structures; flat, strong edge, corner, texture, and weak edge for zeroth and second order ($N = 0$ and 2). (Note: Each weight function is respectively normalized, and Fig. 2 illustrates the detail of the weight function at the strong edge.)

where

$$\mathbf{y} = \left[y_1, y_2, \cdots, y_P\right]^T, \quad \mathbf{b} = \left[\beta_0, \boldsymbol{\beta}_1^T, \cdots, \boldsymbol{\beta}_N^T\right]^T, \tag{8}$$

$$\mathbf{W}_{\mathbf{x}} = \mathrm{diag}\Big[K_{\mathbf{H}_s}(\mathbf{x}_1 - \mathbf{x})K_{h_r}(y_1 - y),\ K_{\mathbf{H}_s}(\mathbf{x}_2 - \mathbf{x})K_{h_r}(y_2 - y),\ \cdots, K_{\mathbf{H}_s}(\mathbf{x}_P - \mathbf{x})K_{h_r}(y_P - y)\Big], \tag{9}$$

and

$$\mathbf{X}_{\mathbf{x}} = \begin{bmatrix} 1 & (\mathbf{x}_1 - \mathbf{x})^T & \mathrm{vech}^T\{(\mathbf{x}_1 - \mathbf{x})(\mathbf{x}_1 - \mathbf{x})^T\} & \cdots \\ 1 & (\mathbf{x}_2 - \mathbf{x})^T & \mathrm{vech}^T\{(\mathbf{x}_2 - \mathbf{x})(\mathbf{x}_2 - \mathbf{x})^T\} & \cdots \\ \vdots & \vdots & \vdots & \vdots \\ 1 & (\mathbf{x}_P - \mathbf{x})^T & \mathrm{vech}^T\{(\mathbf{x}_P - \mathbf{x})(\mathbf{x}_P - \mathbf{x})^T\} & \cdots \end{bmatrix} \tag{10}$$

with "diag" defining a diagonal matrix. Regardless of the regression order (N), since our primary interest is to compute an estimate of the image (pixel values), the necessary computations are limited to the ones that estimate the parameter β_0. Therefore, the weighted least-squares estimation is simplified to

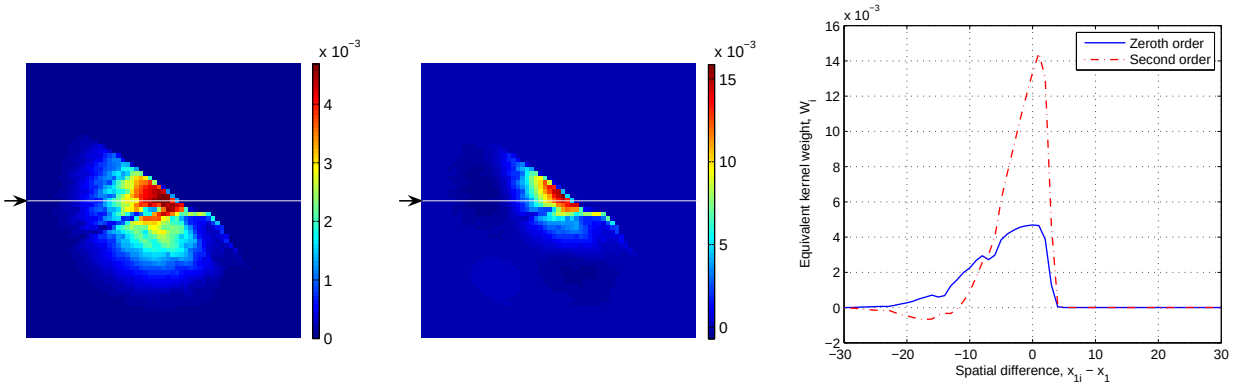
$$\hat{z}(\mathbf{x}) = \hat{\beta}_0 = \mathbf{e}_1^T \left(\mathbf{X}_{\mathbf{x}}^T \mathbf{W}_{\mathbf{x}} \mathbf{X}_{\mathbf{x}}\right)^{-1} \mathbf{X}_{\mathbf{x}}^T \mathbf{W}_{\mathbf{x}}\, \mathbf{y}, \tag{11}$$

where $\mathbf{e}_1$ is a column vector with the first element equal to one, and the rest equal to zero. This estimator can be summarized the form of the weighted linear combination of all the samples using the bilateral "equivalent" weight function W_i as follows:

$$\hat{z}(\mathbf{x}) = \hat{\beta}_0 = \sum_{i=1}^{P} W_i(\mathbf{x}_i - \mathbf{x}, y_i - y, N)\, y_i. \tag{12}$$

For $N = 0$, the estimator (12) becomes a generalization of the Nadaraya-Watson estimator[15] with the bilateral kernel:

$$\hat{z}(\mathbf{x}) = \hat{\beta}_0 = \sum_{i=1}^{P} W_i(\mathbf{x}_i - \mathbf{x}, y_i - y, 0)\, y_i = \frac{\sum_{i=1}^{P} K_{\mathbf{H}_s}(\mathbf{x}_i - \mathbf{x})K_{h_r}(y_i - y)\, y_i}{\sum_{i=1}^{P} K_{\mathbf{H}_s}(\mathbf{x}_i - \mathbf{x})K_{h_r}(y_i - y)}, \tag{13}$$



(a) Zeroth order (b) Second order (c) Horizontal slices of W_i

Figure 2. The footprints of the bilateral equivalent weight function $W_i(\mathbf{x}_i - \mathbf{x}, y_i - y, N)$ at the strong edge for zeroth and second order ($N = 0$ and 2): (a)-(b) the footprints of W_i for zeroth and second order, respectively, (c) the horizontal slices pointed by the arrows of W_i of (a) and (b).

which is nothing but the recently well-studied and popular bilateral filter.[1,2] Fig. 1 illustrates the bilateral equivalent weight function W_i at a variety of image structures for the zeroth and second order cases ($N = 0$ and 2) (Note: Each weight function is respectively normalized), and Fig. 2 illustrates the details of W_i at the strong edge: (a)-(b) the footprints of W_i for zeroth and second order, respectively, (c) the horizontal slices pointed by the arrows of W_i of (a) and (b). The derivations for the bilateral filter (13) and the higher order bilateral filters for $N = 1, 2$ are fully explained in Ref. 12. We note that in general since the pixel value y at an arbitrary position $\mathbf{x}$ might not be available from the given data (e.g. in an interpolation setting), the direct application of (6) is limited to the denoising problem. This limitation, however, can be overcome by using an initial estimate of y by an appropriate interpolation technique (e.g. KR (5)), subsequently followed by the application of data-adaptive kernel regression as elucidated in Ref. 12.

The derivation of the zeroth order bilateral filter (13) above indicates that the bilateral filter only consists of the constant term (β_0), and the consideration of only β_0 is the cause of filtered image signals being piecewise constant. Such signals are often unsuitable for image processing because image signals of interest have complicated contours which include texture and gradation. Higher order bilateral filters relax the piecewise constancy by the choice of $N > 0$. For instance, images estimated by BKR with $N = 1$ and $N = 2$ become piecewise linear and piecewise quadratic, respectively.

As a further extension of the standard bilateral filter, Elad suggested iterative filtering in order to intensify the smoothing effect in Ref. 2. The iterative filtering process is as follows: (i) apply bilateral filter to given noisy data, (ii) apply bilateral filter to the previous estimate, (iii) repeat the step (ii). For $N = 0$, such estimator can be written as

$$\hat{z}^{\ell+1}(\mathbf{x}) = \frac{\sum_{i=1}^{P} K_{\mathbf{H}_s}(\mathbf{x}_i - \mathbf{x}) K_{h_r}\big(\hat{z}^{\ell}(\mathbf{x}_i) - \hat{z}^{\ell}(\mathbf{x})\big)\, \hat{z}^{\ell}(\mathbf{x}_i)}{\sum_{i=1}^{P} K_{\mathbf{H}_s}(\mathbf{x}_i - \mathbf{x}) K_{h_r}\big(\hat{z}^{\ell}(\mathbf{x}_i) - \hat{z}^{\ell}(\mathbf{x})\big)}, \tag{14}$$

where $\hat{z}^{0}(\mathbf{x}_i) = y_i$ and ℓ is the index of iterations. This filtering algorithm is very similar to the Mean-Shift algorithm.[16,17] In the next section, we show a simulation of higher order bilateral filter and its iterative filtering application, and compare to the results of standard (zeroth-order) bilateral filtering.

3. SIMULATION

First, we did denoising simulations by standard and higher order bilateral filters for a variety of SNRs: namely, 10, 15, 20, 25, and 30[dB]. At each noise level, we created 50 noisy images by adding white Gaussian noise with different realizations, and denoised by the standard bilateral filter (13) and the second order bilateral filter. The spatial and radiometric smoothing parameters (h_s and h_r) in each case were optimized by the cross validation

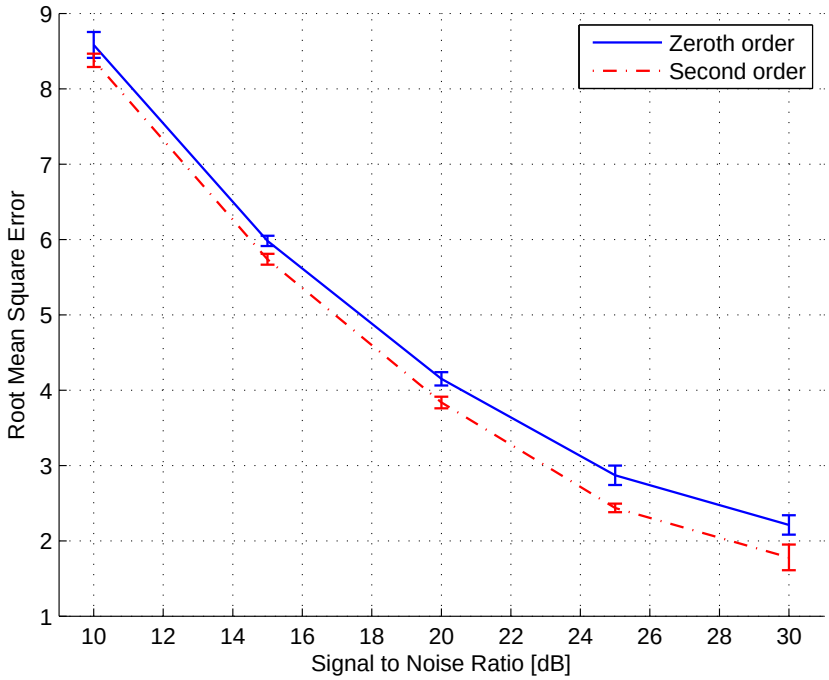


Figure 3. Performance comparison between zeroth and second order bilateral filters by Monte Carlo simulations (50 times). The solid (blue) line and dashed (red) line are mean values and standard deviations of RMSE values in different SNRs by bilateral filter ($N = 0$) and second order bilateral filter ($N = 2$), respectively. The spatial and radiometric smoothing parameters are optimized by the cross validation method.

(CV) method[†].[18] The average root mean square errors (RMSE) are shown in Fig. 3. Numerically, the second order bilateral filter is better than the bilateral filter at all noise levels for this example. Fig. 4 illustrates (a) the original image, (b) a noisy image (SNR = 20[dB]), and (c)-(d) denoised images by the bilateral filter (13) and the second order bilateral filter, respectively. Fig. 5(a) and Fig. 5(b) illustrate absolute residual images of Fig. 4(a) and Fig. 4(b), respectively. It is worth noting that some textures of the original image are visible in Fig. 5(a), and that the residual image in Fig. 5(b) is more noise-like than Fig. 5(a), indicating that the second order bilateral filter was more effective. It is also worth mentioning that in general, the zeroth order bilateral filter will perform better on average for denoising strictly piecewise constant images, but not so for any other general class of images. A further question worth considering is whether a scheme that would change the order N in a spatially varying way would lead to better results for a wider class of images.

Next, the graph illustrated in Fig. 6 shows the behavior of iterative filtering using 50 Monte Carlo simulations in the case of SNR = 20[dB]. The solid (blue) line and the dashed (red) line are average RMSE values by the standard bilateral filter (13) and the second order bilateral filter, respectively, versus the number of iterations. We used the optimized spatial and radiometric smoothing parameters by the cross validation method in the previous simulation. The first filtered images by the bilateral filter and the second order bilateral filter are shown in Fig. 4(a) and Fig. 4(b), respectively, and Fig. 7 shows the filtered images after three and seven iterations. The iterative zeroth order bilateral filter outputs piecewise constant images, and this property can be used for image segmentation, similarly to the output produced by the Mean Shift algorithm. On the other hand, the iterative second order bilateral filter outputs piecewise quadratic images, and the RMSE values for this example become worse as a function of iterations much slower than the iterative zeroth order bilateral filter.

[†]In order to use CV, initial estimates for eliminated pixels are necessary, and we filled up the eliminated pixels by the second order classic kernel regression (5) with the global smoothing parameter $h = 0.4$.

(a) Original image (b) Noisy image (SNR = 20[dB])

(c) Bilateral filter (d) Second order bilateral filter

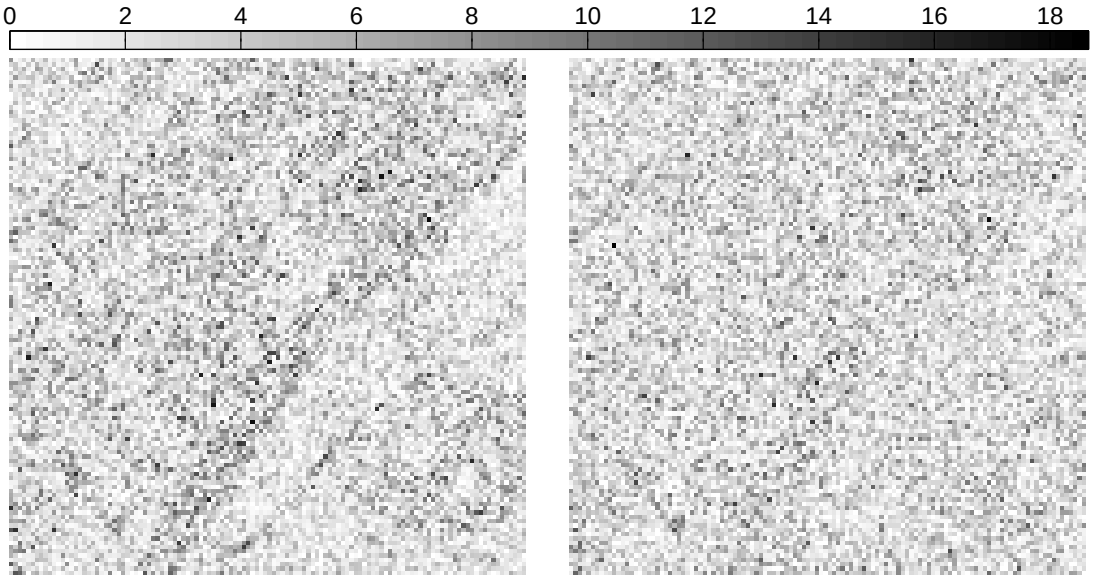
Figure 4. Denoising example of higher order bilateral filter (single application) in SNR = 20[dB]: (a) original image, (b) noisy image (SNR = 20[dB]), and (c)-(d) denoised images by the standard bilateral filter and the second order bilateral filter, respectively. The RMSE values for the denoised images are (c) 4.15 and (d) 3.84.

4. CONCLUSION AND FUTURE WORKS

We verified that the higher order bilateral filter, which was introduced in Ref. 12, is numerically better than the standard bilateral filter for image denoising, and studied its behavior in an iterative setting. In Ref. 19, we showed that the framework of general data-adaptive kernel regression methods, which includes the higher order bilateral filters, can be extended to other image reconstruction problems such as deblurring, and also indicated its applicability to the multi-frame scenario,[12] namely super-resolution.

REFERENCES

1. C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," *Proceeding of the 1998 IEEE International Conference of Compute Vision, Bombay, India* , pp. 836–846, January 1998.
2. M. Elad, "On the origin of the bilateral filter and ways to improve it," *IEEE Transactions on Image Processing* **11**, pp. 1141–1150, October 2002.



(a) Bilateral filter (b) Second order bilateral filter

Figure 5. Absolute residuals images of the standard bilateral filter and the second order bilateral filter: (a) the standard bilateral filter, and (b) the second order bilateral filter.

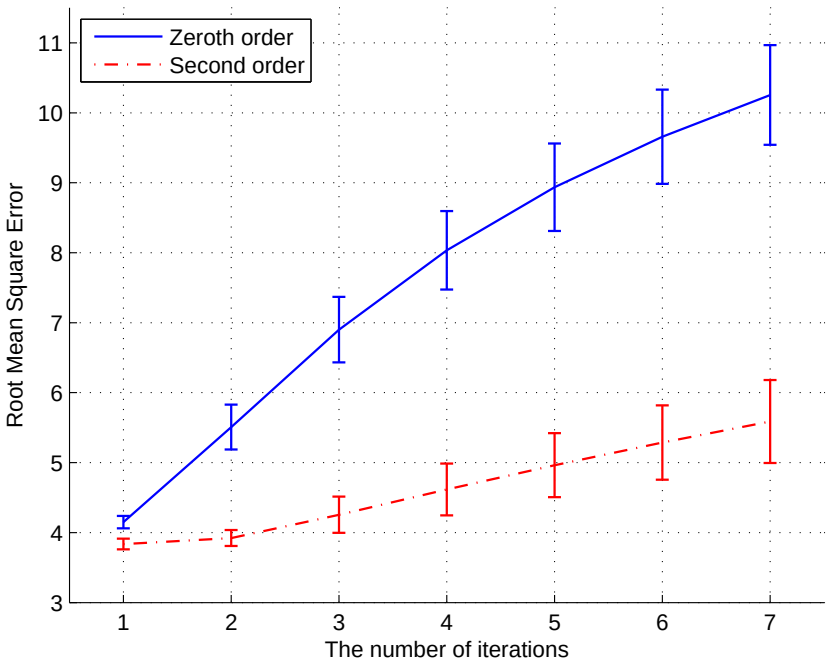


Figure 6. Behavior demonstration of iterative higher order bilateral filters by Monte Carlo simulations (50 times) in SNR = 20[dB]. The solid (blue) line and the dashed (red) line are average RMSE values by bilateral filter ($N = 0$) and second order bilateral filter ($N = 2$), respectively, versus the number of iterations. The spatial and radiometric smoothing parameters are optimized by the cross validation method.

(a) Bilateral filter (3 iterations) (b) Second order bilateral filter (3 iterations)

(c) Bilateral filter (7 iterations) (d) Second order bilateral filter (7 iterations)

Figure 7. Behavior of iterative higher order bilateral filters: (a) and (c) are the images by applying bilateral filter 3 times and 7 times, respectively, and (b) and (d) are the images by applying second order bilateral filter 3 times and 7 times, respectively. The noise level SNR = 20[dB] was used.

3. M. P. Wand and M. C. Jones, *Kernel Smoothing*, Monographs on Statistics and Applied Probability, Chapman and Hall, London; New York, 1995.
4. D. Barash, "A fundamental relationship between bilateral filtering, adaptive smoothing and the nonlinear diffusion equation," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **24**, pp. 844–847, June 2002.
5. D. Barash and D. Comaniciu, "A common framework for nonlinear diffusion, adaptive smoothing, bilateral filtering and mean shift," *Image Vis. Comput.* **22**(1), pp. 73–81, 2004.
6. T. Q. Pham and L. J. van Vliet, "Separable bilateral filtering for fast video preprocessing," *Proceedings of the IEEE International Conference on Multimedia and Expo (ICME) 2005.* , July 2005.
7. W.-C. Kao and Y.-J. Chen, "Multistage bilateral noise filtering and edge detection for color image enhancement," *IEEE Transactions on Consumer Electronics* **51**, pp. 1346–1351, November 2005.

8. S. Morillas, V. Gregori, and A. Sapena, "Fuzzy bilateral filtering for color images," in *Image analysis and recognition, Lecture Notes in Computer Science* **4141**, pp. 138–145, Springer-Verlag, 2006.
9. S. Farsiu, D. Robinson, M. Elad, and P. Milanfar, "Fast and robust multi-frame super-resolution," *IEEE Transactions on Image Processing* **13**, pp. 1327–1344, October 2004.
10. W. Shao and Z. Wei, "Fast and robust filtering-based image magnification," in *Vision Algorithms: Theory and Practice, Image Analysis and Recognition* **4141**, pp. 53–62, 2006.
11. L. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D* **60**, pp. 259–268, November 1992.
12. H. Takeda, S. Farsiu, and P. Milanfar, "Kernel regression for image processing and reconstruction," to appear in IEEE Transactions on Image Processing, February, 2007.
13. N. K. Bose and N. Ahuja, "Superresolution and noise filtering using moving least squares," *IEEE Transactions on Image Processing* **15**, pp. 2239–2248, August 2006.
14. S. M. Kay, *Fundamentals of Statistical Signal Processing - Estimation Theory* -, Signal Processing Series, PTR Prentice-Hall, Englewood Cliffs, N.J., 1993.
15. E. A. Nadaraya, "On estimating regression," *Theory of Probability and its Applications* , pp. 141–142, September 1964.
16. K. Fukunaga and L. D. Hostetler, "The estimation of the gradient of a density function, with applications in pattern recognition," *IEEE Transactions of Information Theory* **21**, pp. 32–40, 1975.
17. D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **24**, pp. 603–619, May 2002.
18. M. Stone, "Cross validatory choice and assessment of statistical predictions," *Journal of the Royal Statistical Society, Series B (Methodological)* **36**(2), pp. 111–147, 1974.
19. H. Takeda, S. Farsiu, and P. Milanfar, "Regularized kernel regression for image deblurring," *Proceedings of the 40th Asilomar Conference on Signals, Systems, and Computers, Pacific Grove, CA* , November 2006.