# Analysis of the Bilateral Filter

Michael Elad
The Computer Science Department, SCCM Program - Stanford University
Email: elad@sccm.stanford.edu

*Abstract*

Effective methods for image denoising are typically based on iterative and locally adaptive algorithms. Recently, an alternative algorithm called 'bilateral filter' was proposed for the same task. This filter was shown to give similar and possibly better results compared to the ones obtained by the best iterative approaches. In this paper we present a relation between the bilateral filter and the Bayesian methodology. Based on this observation we show how the bilateral filter can be improved and extended to treat more general signal characteristics.

## 1. Introduction

Removal of additive noise from a signal is an important problem in signal and image processing [1-7]. This problem is the most simplified reconstruction problem in the wider field of signal restoration [8-9]. Restoring a signal based on corrupted measurements of it is typically done by the Bayesian approach. This approach uses a statistical estimator applied on a Gibbs distribution, resulting with a penalty functional. This functional is minimized by a numerical optimization algorithm that yields the restored signal [8-9].

Noise removal is a practical problem raised in many systems. Apart from the trivial application of removing noise prior to presenting the signal to a human observer, pre-smoothing a signal and noise removal may help to improve the performance for many signal-processing algorithms such as compression, detection, enhancement, recognition, and more. From this aspect, noise removal is appealing both because it relies on a well-established theory, and also because the proposed algorithms in this field are efficient and thus practical.

The more advanced methods for noise removal aim at preserving the signal details while removing the noise. This is achieved by a locally adaptive recovery paradigm. Such methods can be based on Anisotropic Diffusion (AD) [1-4], Weighted Least Squares (WLS) [5], or Robust Estimation (RE) [6-7]. The Mumford-Shah functional is a different yet resembling approach toward the same denoising task [10]. All these methods share the fact that local relations between the samples dictate the final result, and therefore, all these methods resort to an iterative algorithm. There is a solid theoretical bridge between these methods as well as to the Line-Process approach [11-12].

Recently, Tomasi and Manduchi proposed an alternative non-iterative bilateral filter for removing noise from images [13]. This filter is a weighted average of the local neighborhood samples, where the weights are computed based on temporal (or spatial in case of images) and radiometric distances between the center sample and its neighbors. This filter was shown to give similar and possibly better results compared to those obtained by the previously mentioned iterative approaches. However, The bilateral filter was proposed in [13] as an intuitive tool. In this paper we explore its theoretical relation to the AD, the WLS, and the RE techniques, and show that the bilateral filter also emerges from the Bayesian methodology, using a novel penalty functional. For this functional, we show that a single iteration of the Jacoby algorithm yields the bilateral filter. Based on this observation, we also show how the bilateral filter can be improved to further speed-up its smoothing operation, and show how this filter can be extended to treat piece-wise linear signals. Also, we should mention that the bilateral filter could be extended to treat more general reconstruction problems such as image restoration, image scaling, super-resolution and more.

In the next section we shortly describe the bilateral filter. Section 3 describes the AD, the WLS, and the RE methods. In Section 4 we propose a novel penalty term, strongly related to the one in Section 3. We show how this new penalty term yields the bilateral filter. Section 5 discusses several improvements to the bilateral filter based on our new model. In Section 6 we compare the various methods discussed in this paper for simple 1D signals. We summarize this paper in Section 7. A wider and more detailed description of the results given here can be found in [16].

## 2. Noise Suppression Via the Bilateral Filter

We start our discussion with a presentation of the bilateral filter as proposed originally by Tomasi and Manduchi [13]. In order to simplify the notations we stick to the 1D case throughout this paper, though all derivations apply to the 2D case just as well.

An unknown signal $\underline{X}$ represented as a vector goes through a degradation stage in which a zero-mean white Gaussian noise $\underline{V}$ is added to it. The result is the corrupted signal $\underline{Y}$ given by

$$\underline{Y} = \underline{X} + \underline{V} . \tag{2.1}$$

Our task is to remove this noise and restore $\underline{X}$, given the degraded signal $\underline{Y}$. The bilateral filter suggests a weighted average of pixels in the given image $\underline{Y}$

$$\hat{X}[k] = \frac{\sum_{n=-N}^{N} W[k,n] Y[k-n]}{\sum_{n=-N}^{N} W[k,n]} \quad . \tag{2.2}$$

This equation is simply a normalized weighted average of a neighborhood of $[2N+1]$ samples around the $k^{th}$ sample. The weights W[k,n] are computed based on the content of the neighborhood. For the center sample X[k], the weight W[k,n] is computed by multiplying the following two factors:

$$\begin{aligned} W_S[k,n] &= \exp\left\{-\frac{d^2\{[k],[k-n]\}}{2\sigma_S^2}\right\} = \exp\left\{-\frac{n^2}{2\sigma_S^2}\right\} \\ W_R[k,n] &= \exp\left\{-\frac{d^2\{Y[k],Y[k-n]\}}{2\sigma_R^2}\right\} = \\ &= \exp\left\{-\frac{[Y[k]-Y[k-n]]^2}{2\sigma_R^2}\right\}. \end{aligned} \tag{2.3}$$

The final weight is obtained by multiplying the two

$$W[k,n] = W_S[k,n] \cdot W_R[k,n] \, . \tag{2.4}$$

The weight includes two ingredients – temporal (spatial in case of images) and radiometric weights. The first weight measures the geometric distance between the center sample [k] and the [k-n] sample, and Euclidean metric is applied here. This way, close-by samples influence the final result more than distant ones. The second weight measures the radiometric distance between the values of the center sample Y[k] and the [k-n] sample, and again, Euclidean metric is chosen. Therefore, samples with close-by values tend to influence the final result more than those having distant value. Of-course, for both weights we are free to adopt any other reasonable metric. Also, instead of using the Gaussian function, other symmetric and smoothly decaying functions can be used.

## 3. Anisotropic Diffusion, WLS and RE

For the same denoising problem described above, a known approach is to define a penalty functional that best represents our requirements from the unknown $\underline{X}$. We want the result to be as close as possible to the measured signal $\underline{Y}$ while being smooth. Smoothness should be forced in a temporally (spatially) dependent manner in order not to suppress edges in the signal $\underline{X}$. Thus, one either uses weighted least squares (WLS) [5]

$$\begin{aligned} \varepsilon_{WLS}\{\underline{X}\} &= \frac{1}{2}[\underline{X}-\underline{Y}]^T[\underline{X}-\underline{Y}] + \\ &+ \frac{\lambda}{2}[\underline{X}-\mathbf{D}\underline{X}]^T \mathbf{W}(\underline{Y})[\underline{X}-\mathbf{D}\underline{X}] \end{aligned} \tag{3.1}$$

or a robust estimation technique (RE), using an 'M-function' denoted as $\rho(\alpha)$ [6-7]

$$\varepsilon_{RE}\{\underline{X}\} = \frac{1}{2}[\underline{X}-\underline{Y}]^T[\underline{X}-\underline{Y}] + \frac{\lambda}{2}\rho\{\underline{X}-\mathbf{D}\underline{X}\} \, . \tag{3.2}$$

The matrix **D** stands for a one-sample shift to the right (towards the origin) operation. Thus, the term ($\underline{X}$-**D**$\underline{X}$) is a discrete approximation of a backward first derivative. The matrix **W** is a diagonal matrix that weights the local gradients. In the RE, $\rho(\alpha)$ is symmetric non-negative function that penalizes gradient values. The choice $\rho(\alpha) = 0.5\alpha^2$ gives the trivial LS approach. Both these penalty functionals can be shown to emerge from the Bayesian framework and represent the MAP estimation [5-9]. In both cases, an iterative algorithm is typically required in order to find the signal $\underline{X}$ that minimizes the functionals [5-7]. A natural choice is the Steepest Descent (SD) algorithm due to its simplicity [14]. This algorithm requires the computation of the first derivative of the functionals

$$\frac{\partial\varepsilon_{WLS}\{\underline{X}\}}{\partial\underline{X}} = [\underline{X}-\underline{Y}] + \lambda[\mathbf{I}-\mathbf{D}]^T \mathbf{W}(\underline{Y})[\mathbf{I}-\mathbf{D}]\underline{X}, \tag{3.3}$$

$$\frac{\partial\varepsilon_{RE}\{\underline{X}\}}{\partial\underline{X}} = [\underline{X}-\underline{Y}] + \lambda[\mathbf{I}-\mathbf{D}]^T \rho'\{[\mathbf{I}-\mathbf{D}]\underline{X}\} \, . \tag{3.4}$$

Using $\underline{Y}$ as initialization gives

$$\begin{aligned} \hat{\underline{X}}_1^{WLS} &= \hat{\underline{X}}_0^{WLS} - \mu \left.\frac{\partial\varepsilon_{WLS}\{\underline{X}\}}{\partial\underline{X}}\right|_{\underline{X}=\hat{\underline{X}}_0^{WLS}} = \\ &= \hat{\underline{X}}_0^{WLS} - \mu\Big[\Big(\hat{\underline{X}}_0^{WLS} - \underline{Y}\Big) + \\ &+ \lambda(\mathbf{I}-\mathbf{D})^T \mathbf{W}(\underline{Y})(\mathbf{I}-\mathbf{D})\hat{\underline{X}}_0^{WLS}\Big] = \\ &= \underline{Y} - \mu\lambda(\mathbf{I}-\mathbf{D})^T \mathbf{W}(\underline{Y})(\mathbf{I}-\mathbf{D})\underline{Y} \end{aligned} \tag{3.5}$$

or

$$\begin{aligned} \hat{\underline{X}}_1^{RE} &= \hat{\underline{X}}_0^{RE} - \mu \left.\frac{\partial\varepsilon_{RE}\{\underline{X}\}}{\partial\underline{X}}\right|_{\underline{X}=\hat{\underline{X}}_0^{RE}} = \\ &= \hat{\underline{X}}_0^{RE} - \mu\Big[\Big(\hat{\underline{X}}_0^{RE} - \underline{Y}\Big) + \\ &+ \lambda(\mathbf{I}-\mathbf{D})^T \rho'\Big\{(\mathbf{I}-\mathbf{D})\hat{\underline{X}}_0^{RE}\Big\}\Big] = \\ &= \underline{Y} - \mu\lambda(\mathbf{I}-\mathbf{D})^T \rho'\{(\mathbf{I}-\mathbf{D})\underline{Y}\}. \end{aligned} \tag{3.6}$$

Looking at both iterative procedures, we see that they will produce the same solution after the first iteration if

$$\begin{aligned} &\forall\underline{Y},\ \mathbf{W}(\underline{Y})(\mathbf{I}-\mathbf{D})\underline{Y} = \rho'\{(\mathbf{I}-\mathbf{D})\underline{Y}\} \\ &\Rightarrow\ \mathbf{W}(\underline{Y}) = \frac{\rho'\{(\mathbf{I}-\mathbf{D})\underline{Y}\}}{(\mathbf{I}-\mathbf{D})\underline{Y}} \end{aligned} \quad , \tag{3.7}$$

where the above division is applied entry-by-entry. Black, Sapiro, Marimont and Heeger linked between the anisotropic diffusion and the robust estimator and obtained a similar formula [11].

The Anisotropic Diffusion (AD) is different in the sense that it uses the continuum to represent its behavior [1-4]. However, since eventually we work on a discrete signal, we discretize the propagation equations and get a similar equation to the one shown for the RE method [11].

The AD, the WLS and the RE algorithms are based on a solid theory of statistical estimators and regularization theory [1-7]. The bilateral filter, on the other hand, is an ad-hoc filter without theoretic background, and nevertheless with impressive results.

## 4. Derivation of the Bilateral Filter

We propose the following new penalty functional for the unknown signal $\underline{X}$

$$\varepsilon\{\underline{X}\} = \frac{1}{2}[\underline{X}-\underline{Y}]^T[\underline{X}-\underline{Y}] + \frac{\lambda}{2}\sum_{n=1}^{N}\left[\underline{X}-\mathbf{D}^n\underline{X}\right]^T \mathbf{W}(\underline{Y},n)\left[\underline{X}-\mathbf{D}^n\underline{X}\right]. \quad (4.1)$$

When the matrix **D** is raised to the power n it implies a shift right of n samples. Thus, as opposed to the previous smoothness terms, the difference between this functional and the one presented in Equation (3.1) is the use of several scales of derivatives, all applied directly on the unknown image. Taking the first derivative of Equation (4.1) with respect to the unknown $\underline{X}$ we get the following gradient vector

$$\begin{aligned}\frac{\partial\varepsilon\{\underline{X}\}}{\partial\underline{X}} &= [\underline{X}-\underline{Y}] + \lambda\sum_{n=1}^{N}\left(\mathbf{I}-\mathbf{D}^n\right)^T\mathbf{W}(\underline{Y},n)\left(\mathbf{I}-\mathbf{D}^n\right)\underline{X} \\ &= \left[\mathbf{I}+\lambda\sum_{n=1}^{N}\left(\mathbf{I}-\mathbf{D}^n\right)^T\mathbf{W}(\underline{Y},n)\left(\mathbf{I}-\mathbf{D}^n\right)\right]\underline{X}-\underline{Y} = \\ &= \left[\mathbf{I}+\lambda\sum_{n=1}^{N}\left(\mathbf{I}-\mathbf{D}^{-n}\right)\mathbf{W}(\underline{Y},n)\left(\mathbf{I}-\mathbf{D}^n\right)\right]\underline{X}-\underline{Y},\end{aligned} \quad (4.2)$$

If we assume again a single iteration of the SD algorithm applied with $\underline{Y}$ as the initialization, we get

$$\begin{aligned}\underline{\hat{X}}_1 &= \underline{\hat{X}}_0 - \mu\left[\mathbf{I}+\lambda\sum_{n=1}^{N}\left(\mathbf{I}-\mathbf{D}^{-n}\right)\mathbf{W}(\underline{Y},n)\left(\mathbf{I}-\mathbf{D}^n\right)\right]\underline{X}_0 \\ &+\mu\underline{Y} = \left[\mathbf{I}-\mu\lambda\sum_{n=1}^{N}\left(\mathbf{I}-\mathbf{D}^{-n}\right)\mathbf{W}(\underline{Y},n)\left(\mathbf{I}-\mathbf{D}^n\right)\right]\underline{Y}.\end{aligned} \quad (4.3)$$

Speeding-up the above iteration can be done using locally adaptive step-size, obtained by the Jacoby algorithm [14]. The Hessian of our functional is the following matrix

$$\frac{\partial^2\varepsilon\{\underline{X}\}}{\partial\underline{X}^2} = \mathbf{H}(\underline{Y}) = \mathbf{I}+\lambda\sum_{n=1}^{N}\left(\mathbf{I}-\mathbf{D}^{-n}\right)\mathbf{W}(\underline{Y},n)\left(\mathbf{I}-\mathbf{D}^n\right). \quad (4.4)$$

From this matrix we need to extract the main diagonal, which contains real and positive values. We define a step-size matrix **M**, which extends the notion of the previously used μ by $\mathbf{M}(\underline{Y}) = \left[\xi\mathbf{I}+\mathrm{diag}\{\mathbf{H}(\underline{Y})\}\right]^{-1}$. The additional term $\xi\mathbf{I}$ relaxes the step-size matrix and ensures stability. Thus, the Jacoby iteration is

$$\begin{aligned}\underline{\hat{X}}_1 &= \underline{Y} - \lambda\mathbf{M}(\underline{Y})\sum_{n=1}^{N}\left(\mathbf{I}-\mathbf{D}^{-n}\right)\mathbf{W}(\underline{Y},n)\left(\mathbf{I}-\mathbf{D}^n\right)\underline{Y} = \\ &= \left[\mathbf{I}-\lambda\mathbf{M}(\underline{Y})\sum_{n=1}^{N}\left(\mathbf{I}-\mathbf{D}^{-n}\right)\mathbf{W}(\underline{Y},n)\left(\mathbf{I}-\mathbf{D}^n\right)\right]\underline{Y}.\end{aligned} \quad (4.5)$$

Choosing the weights can be done using the Equation (3.7), but they should also reflect our decreased confidence in the smoothness penalty term as n grows towards N. Thus, a reasonable choice is

$$\mathbf{W}(\underline{Y},n) = \frac{\rho'\left\{\left(\mathbf{I}-\mathbf{D}^n\right)\underline{Y}\right\}}{\left(\mathbf{I}-\mathbf{D}^n\right)\underline{Y}}\cdot V(n) \quad (4.6)$$

for some non-negative symmetric and monotonically decreasing function V(n) (e.g. $V(n)=\alpha^n,\ 0<\alpha<1$). We can refer to the above expression as a time varying convolution of the form

$$\hat{X}_1[k] = \sum_{n=-N}^{N} f[l,k]\cdot Y[k-l]. \quad (4.7)$$

After several (tedious) algebraic steps we obtain

$$f[l,k] = \begin{cases} \dfrac{\lambda V(l)\cdot\dfrac{\rho'\{Y[k]-Y[k-l]\}}{(Y[k]-Y[k-l])}}{\xi+1+\lambda\sum\limits_{n=-N}^{N}\dfrac{V(n)\cdot\rho'\{Y[k]-Y[k-n]\}}{Y[k]-Y[k-n]}} & \begin{bmatrix} N\le l\le N, \\ l\ne 0\end{bmatrix} \\[2ex] \dfrac{\xi+1}{\xi+1+\lambda\sum\limits_{n=-N}^{N}\dfrac{V(n)\cdot\rho'\{Y[k]-Y[k-n]\}}{Y[k]-Y[k-n]}} & [l=0]. \end{cases} \quad (4.8)$$

We see that the sum of all coefficients is 1, as should be in the bilateral filter. If the M-function $\rho(\alpha)$ is symmetric and monotonic non-decreasing (from 0 to $\infty$), all the filter coefficients are non-negative. The coefficient $f[l,k]$ represents the weight according to which $Y[k-l]$ contributes to the evaluation of the restored pixel X[k]. This coefficient includes two parts: the spatial weight $V(l)$ and the radiometric weight given by $\rho'\{Y[k]-Y[k-l]\}/(Y[k]-Y[k-l])$. These two parts are the same as described in section 2 for the Tomasi Manduchi bilateral filter. Thus, we get the same filter as in the bilateral filter in [14] if we choose

$$V(l) = \exp\left\{-\frac{l^2}{2\sigma_S^2}\right\} \quad (4.9)$$

$$\frac{\rho'\{Y[k]-Y[k-l]\}}{(Y[k]-Y[k-l])} = \exp\left\{-\frac{[Y[k]-Y[k-n]]^2}{2\sigma_R^2}\right\}$$

$$\Rightarrow \rho(\alpha) = -\sigma_R^2\exp\left\{-\frac{\alpha^2}{2\sigma_R^2}\right\}$$

## 5. Improvements of the Bilateral Filter

The bilateral filter can be speeded-up in one of two methods, and combination of them. Given a general quadratic penalty function of the form

$$\varepsilon\{\underline{X}\} = \sum_{j=1}^{J}\left[\frac{1}{2}\underline{X}^T Q_j\underline{X} - \underline{P}_j^T\underline{X} + C_j\right], \quad (5.1)$$

the SD iteration reads

$$\underline{\hat{X}}_1 = \underline{\hat{X}}_0 - \left.\frac{\partial \varepsilon\{\underline{X}\}}{\partial \underline{X}}\right|_{\underline{\hat{X}}_0} = \underline{\hat{X}}_0 - \mu \sum_{j=1}^{J} \left[ Q_j \underline{\hat{X}}_0 - \underline{P}_j \right]. \quad (5.2)$$

One way to speed the SD convergence is the Gauss-Siedel approach [15-16], where the samples of $\underline{\hat{X}}_1$ are computed sequentially from $\underline{\hat{X}}_1[1]$ to $\underline{\hat{X}}_1[L]$, and for the calculation of $\underline{\hat{X}}_1[k]$, instead of using only $\underline{\hat{X}}_0$ values, updated values of $\underline{\hat{X}}_1$ are used. This 'bootstrap' method is known to be stable and converge to the same global minimum point of the penalty function given in Equation (5.1) [15-16]. A more systematic way to describe this process is via the decomposition of the Hessian to the upper-triangle, lower-triangle, and diagonal parts.

A different alternative for speeding the bilateral filter is to exploit the fact that the gradient is naturally sliced into several parts. Returning to Equation (5.2) we can update the output after every item from the summation, using it to compute the next gradient terms. Thus, the final solution is closer to the global minimum point of the penalty function in Equation (5.1). Note that by applying J sub-iterations, the computational load is similar to the one required with J iterations of the WLS/RE methods. However, the results are expected to be different, since by applying different kinds of derivatives (due to the use of different neighbors) we get stronger smoothing effect.

When treating non-piece-wise constant signals, the bilateral filter is not expected to perform well, and this could be seen since the penalty functional in Equation (4.1) is designed for piece-wise constant signals. Instead, we may propose a different penalty

$$\varepsilon\{\underline{X}\} = \frac{1}{2}[\underline{X}-\underline{Y}]^T[\underline{X}-\underline{Y}] + \\ + \frac{\lambda}{2}\sum_{n=1}^{N}\left[\underline{X} - \frac{\mathbf{D}^n\underline{X}+\mathbf{D}^{-n}\underline{X}}{2}\right]^T \mathbf{W}(\underline{Y},n)\left[\underline{X} - \frac{\mathbf{D}^n\underline{X}+\mathbf{D}^{-n}\underline{X}}{2}\right]. \quad (5.3)$$

Following the same analysis as before, one may show that there is an effective spatially varying filter (see [18]).

## 6. Simulations

Figure 1 presents a piece-wise constant test image ($\underline{X}$) and its noisy version ($\underline{Y}$). Figure 2 (a and b) shows the results obtained by the WLS and the RE methods. The WLS was applied with weights computed via the assumption $\rho(\alpha)=|\alpha|$, choosing $\lambda=1$, and applying 50 SD iterations. Similarly, the RE used $\rho(\alpha)=|\alpha|$, $\lambda=1$ and 50 iterations. We measure performance by computing the MSE gain, defined as the ratio between the Mean-Square-Error (MSE) before and after the filtering. The obtained MSE gain in the WLS method is 3.90. The MSE gain for the RE is 10.99. Figure 2 (c and d) shows the result of the bilateral filter with weights as in Equations (4.10) and (4.11) with $\lambda=1, \sigma_S=2.5, \sigma_R=0.5, N=6$. A single application of this filter gave an MSE gain of 23.50. The result after 10 iterations of the bilateral filter with an MSE gain of 318.90 is shown in Figure 2 (d). Applying the Gauss-Siedel with the same parameters (and thus, having the exact same complexity) we got an MSE gain of 39.44. Applying the second speed-up approach with the sliced gradient terms we got an MSE gain of 197.26.

Figure 3 (a) shows a piece-wise linear image $\underline{X}$ and its noisy version Y (b). An attempt to recover this signal using the regular bilateral gave an MSE gain of 1.53 (c). A piece-wise-linear bilateral filter as proposed above gave an MSE gain of 12.91 (d).

As a final point in this section, we consider the continuity of the filter coefficients in the RE, and the fact that it is not so for the bilateral filter. Figure 4 shows a noisy piece-wise image (checkerboard). This signal was filtered by the RE (same parameters as before and using 1500 iterations) resulting with an MSE gain of 2.32 (c). A single iteration of the bilateral filter with the parameters ($\lambda=1, \sigma_S=5, \sigma_R=0.5, N=6$) gave an MSE gain of 19.97 (d). The main reason for the much better performance with the bilateral is its ability to create a local filter that has a non-connected structure.

## 7. Summary

In this paper we proposed a theory for explaining the origin of the bilateral filter, and shown that the Bayesian approach is also in the core of the bilateral filter, just as it has been for the AD, WLS and the RE. We have also shown how this new insight can serve for improving the bilateral filter and extend its use for other applications.

## References


[1] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion", IEEE Transaction on PAMI-Vol. 12, No. 7, pp.629-639, 1990.

[2] N. Sochen, R. Kimmel, and R. Malladi, "A geometrical framework for low level vision", IEEE Trans. on Image Processing, Vol. 17, pp. 310-318, 1998.

[3] J. Weickert, Anisotropic diffusion in image processing, ECMI Series, Teubner, Stuttgart, 1998.

[4] G. Sapiro and D. L. Ringach, "Anisotropic diffusion of color images", In Proc. SPIE, 382-2657:471, 1996.

[5] R. L. Lagendijk, J. Biemond, and D. E. Boekee, "Regularized iterative image restoration with ringing reduction," IEEE Trans. Acoustics, Speech, Signal Processing, Vol. 36, pp. 1874-1887, December 1988.

[6] M. E. Zervakis, Nonlinear Image Restoration Techniques. Ph.D. Thesis, University of Toronto, 1990.

[7] M.J. Black and G. Sapiro, "Edges as outliers: anisotropic smoothing using local image statistics", Scale-Space '99. Springer-Verlag, pp. 259-270, Berlin, Germany, 1999.

[8] R. Lagendijk and J. Biemond, Iterative identification and restoration of images, Kluwer Academic, Boston, 1991.

[9] Anil K. Jain, Fundamentals of Digital Image Processing. Prentice Hall, Englewood Cliffs, NJ, 1989.

[10] D. Mumford and J. Shah, "Optimal approximations by piecewise smooth functions and associated variational problems", Comm. Pure Applied Math., pp. 577-685, 1989.

[11] M. J. Black, G. Sapiro, D. H. Marimont, and D. Heeger, "Robust Anisotropic Diffusion", IEEE Transaction on Image Processing, Vol. 7, pp. 421-432, 1998.

[12] M. Black and A. Rangarajan, "On the unification of line processes, outlier rejection, and robust statistics with applications in early vision", International Journal of Computer Vision, Vol. 19, pp. 57-92, 1996.

[13] C. Tomasi, R. Manduchi, "Bilateral filtering for gray and color images", Sixth International Conference on Computer Vision, pp. 839-46. New Delhi, India, 1998.

[14] D. Bertsekas, Nonlinear Programming, Athena, 1995.

[15] J. Weickert, B.M. T.H. Romeny, and M.A. Viergever, "Efficient and reliable schemes for nonlinear diffusion filtering", IEEE Trans. Image Processing, Vol. 7, pp. 398-410, March 1998.

[16] M. Elad, "On the Bilateral Filter and Ways to Improve It", Accepted to the IEEE Transactions On Image Processing on October 2001.

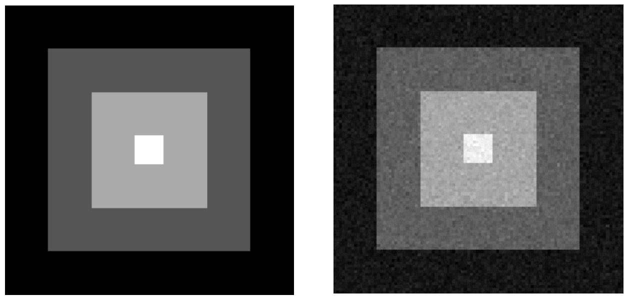
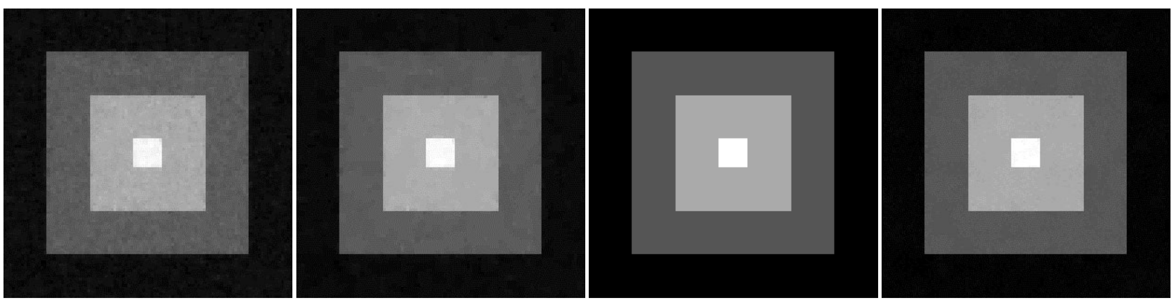
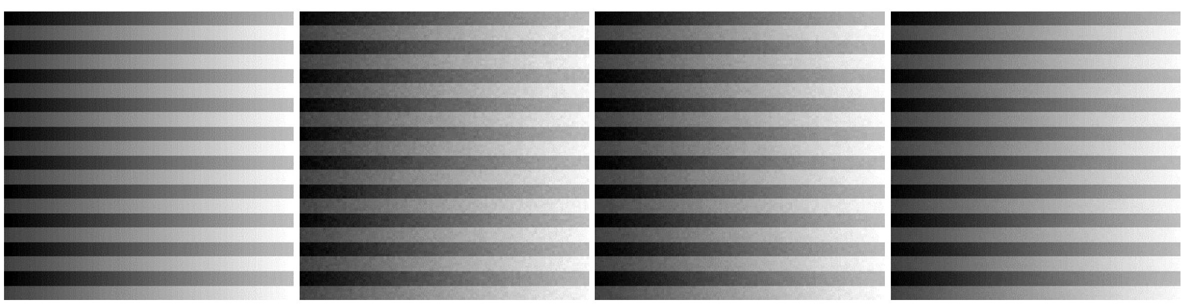


Figure 1 – A piece-wise constant test image (left) and its noisy version (right).



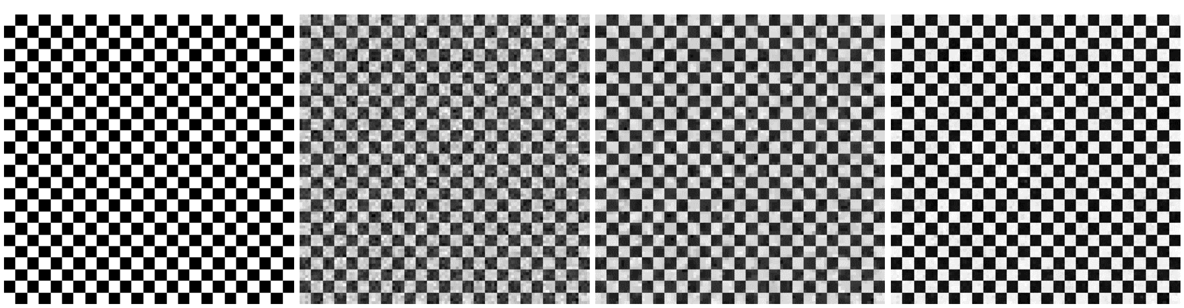
(a) (b) (c) (d)

Figure 2 – (a) The WLS with 50 iterations (MSE gain=3.90); (b) the RE with 50 iterations (MSE gain=10.99); (c) The bilateral filter result after one iteration (MSE gain=23.50); and (d) after 10 iterations (MSE gain=318.90).



(a) (b) (c) (d)

Figure 3 – (a) A piece-wise linear test signal; (b) Its noisy version; (c) The regular bilateral; and (d) The piece-wise-linear bilateral.



(a) (b) (c) (d)

Figure 4 – (a) A piece-wise constant test image; (b) Its noisy version; (c) RE with 1500 iterations; and (d) Bilateral filter.