A NEW APPROACH FOR DENOISING IMAGES BASED ON WEIGHTS OPTIMIZATION

Qiyu Jin[1,2], Ion Grama[1,2] and Quansheng Liu[1,2]
[1]*UMR 6205, Laboratoire de Mathmatiques de Bretagne Atlantique, Université de Bretagne-Sud, Campus de Tohanic, BP 573, 56017 Vannes, France*
[2]*Université Européenne de Bretagne, Rennes, France*

Keywords: Non-local Means, Image Denoising, Optimization of Weights, Oracle, Statistical Estimation.

Abstract: We propose a new algorithm to restore an image contaminated by the Gaussian white noise. Our approach is based on the weighted average of the observations in a neighborhood as in the case of the Non-Local Means Filter. But in contrast to the Non-Local Means Filter, we choose the weights by minimizing a tight upper bound of the Mean Square Error. Our theoretical results show that some "oracle" weights defined by a triangular kernel are optimal. To construct a computable filter the "oracle" weights are replaced by some estimates. The implementation of the proposed algorithm is straightforward. The simulations show that our approach is very competitive.

1 INTRODUCTION

We deal with the additive Gaussian noise model:

$$Y(x) = f(x) + \varepsilon(x),\ x \in \mathbf{I}, \tag{1}$$

where $\mathbf{I}$ is the uniform $N \times N$ grid of pixels on the unit square, $Y = (Y(x))_{x\in\mathbf{I}}$ is the observed image brightness, $f : [0,1]^2 \to \mathbf{R}_+$ is an unknown target regression function and $\varepsilon = (\varepsilon(x))_{x\in\mathbf{I}}$ are independent and identically distributed (i.i.d.) Gaussian random variables with mean 0 and standard deviation $\sigma > 0$.

Important denoising techniques for the model (1) have been developed in recent years. A very significant step in these developments was the introduction of the Non-Local Means Filter by (Buades et al., 2005). For closely related works, see for example (Polzehl and Spokoiny, 2006; Kervrann and Boulanger, 2008; Buades et al., 2009; Katkovnik et al., 2010; Lou et al., 2010).

The basic idea of the filters by weighted means is to estimate the unknown image $f(x_0)$ by a weighted average of observations $Y(x)$ of the form

$$\widetilde{f}_w(x_0) = \sum_{x\in\mathbf{U_{x_0,h}}} w(x)Y(x), \tag{2}$$

where for each x_0 and $h > 0$, $\mathbf{U}_{x_0,h}$ denotes a square window with center x_0 and width $2h$, $w(x)$ are some non-negative weights satisfying $\sum_{x\in\mathbf{U}_{x_0,h}} w(x) = 1$. The choice of the weights $w(x)$ are usually based on two criteria: a spatial criterion so that $w(x)$ is a decreasing function of the distance between x and x_0, and a similarity criterion so that $w(x)$ is also a decreasing function of the brightness difference $|Y(x) - Y(x_0)|$ (see e.g. (Yaroslavsky, 1985; Tomasi and Manduchi, 1998)), which measures the similarity between the pixels x and x_0. In the Non-Local Means Filter, $h > 0$ can be chosen relatively large, and the weights $w(x)$ are calculated according to the similarity between data patches $\mathbf{Y}_{x,\eta} = (Y(y) : y \in \mathbf{U}_{x,\eta})$ (identified as a vector whose composants are ordered lexicographically) and $\mathbf{Y}_{x_0,\eta} = (Y(y) : y \in \mathbf{U}_{x_0,\eta})$, instead of the similarity between just the pixels x and x_0. Here $\eta > 0$ is the size parameter of data patches.

In this paper we address the problem of choosing the weights w in (2) in some optimal way. Generally, the weights w are defined through some priory fixed kernels, often the Gaussian one. The important problem of the choice of the kernel has not been addressed so far. Although the choice of the Gaussian kernel yields good numerical performance, there is no particular reason to restrict ourselves to this kernel. Our theoretical results and simulations show that another kernel is preferred; this kernel leads to us an improved Non-Local Means Filter which also has the advantage that it is parameter free in the sense that it automatically calculates the bandwidth of the smoothing kernel.

Our main idea is to produce a very tight upper

Jin Q., Grama I. and Liu Q..
A NEW APPROACH FOR DENOISING IMAGES BASED ON WEIGHTS OPTIMIZATION.
DOI: 10.5220/0003846001120117
In *Proceedings of the International Conference on Computer Vision Theory and Applications* (VISAPP-2012), pages 112-117
ISBN: 978-989-8565-03-7

bound of the Mean Square Error

$$R\left(\widetilde{f}_w(x_0)\right) = \mathbb{E}\left(\widetilde{f}_w(x_0) - f(x_0)\right)^2$$

in terms of the bias and variance and to minimize this upper bound in w under the constraints $w \geq 0$ and $\sum_{x\in \mathbf{U}_{x_0,h}} w(x) = 1$. We first obtain an explicit formula for the optimal weights w^* in terms of the unknown function f. In order to get a computable filter, we estimate w^* by some adaptive weights $\widehat{w}$ based on data patches from the observed image Y. We thus obtain a new filter, which we call *Optimal Weights Filter*. Numerical results show that the new filter outperforms the typical Non-Local Means Filter, thus giving a practical justification that the optimal choice of the kernel improves the denoising quality.

We would like to point out that related optimization problems for non parametric signal and density recovering have been proposed earlier in (Sacks and Ylvisaker, 1978; Nazin et al., 2008). In these papers the weights are optimized over a given class of regular functions and thus depend only on some parameters of the class. The novelty of our work is to deal with optimal weights depending on the image f at hand. Results of this type are related to the "oracle" concept developed in (Donoho and Johnstone, 1994).

2 OPTIMAL WEIGHTS FILTER

In this section, we present our new filter called Optimal Weights Filter, and explain the idea behind its construction.

We begin with some mathematical notations that will be used throughout the paper. For a vector $x = (x_1,...,x_d) \in \mathbf{R}^d$, we denote by $\|x\|_2 = \left(\sum_{i=1}^d x_i^2\right)$ its Euclidean norm and by $\|x\|_\infty = \max_{1\leq i\leq d}|x_i|$ its supremum norm. The cardinality of a set $\mathbf{A}$ is denoted by card $\mathbf{A}$. For a positive integer N the uniform $N\times N$ grid on the unit square is defined by

$$\mathbf{I} = \left\{\frac{1}{N}, \frac{2}{N}, \cdots, \frac{N-1}{N}, 1\right\}^2.$$

Each element x of the grid $\mathbf{I}$ will be called pixel. The number of pixels is $n = N^2$. For any pixel $x_0 \in \mathbf{I}$ and a given $h > 0$, the square window of pixels $\mathbf{U}_{x_0,h} = \{x \in \mathbf{I} : \|x - x_0\|_\infty \leq h\}$ will be called *search window* at x_0. We naturally take h as a multiple of $\frac{1}{N}$ ($h = \frac{k}{N}$ for some $k \in \{1,2,\cdots,N\}$). The size of the square search window $\mathbf{U}_{x_0,h}$ is the positive integer number

$$M = (2Nh+1)^2 = \text{card}\ \mathbf{U}_{x_0,h}. \tag{3}$$

For any pixel $x \in \mathbf{U}_{x_0,h}$ and a given $\eta > 0$ a second square window of pixels $\mathbf{U}_{x,\eta}$ will be called *patch* at x. Like h, the parameter η is also taken as a multiple of $\frac{1}{N}$. The size of the patch $\mathbf{U}_{x,\eta}$ is the positive integer

$$m = (2N\eta+1)^2 = \text{card}\ \mathbf{U}_{x_0,\eta}. \tag{4}$$

The vector $\mathbf{Y}_{x,\eta} = (Y(y))_{y\in\mathbf{U}_{x,\eta}}$ formed by the values of the observed noisy image in the patch $\mathbf{U}_{x,\eta}$ in the lexicographical order will be called *data patch (or similarity patch)* at $x \in \mathbf{U}_{x_0,h}$. Finally, the positive part of a real number a is denoted by a^+: $a^+ = a$ if $a \geq 0$ and $a^+ = 0$ if $a < 0$.

Let $h > 0$ be fixed. For any pixel $x_0 \in \mathbf{I}$ consider a family of weighted estimates $\widetilde{f}_{h,w}(x_0)$ of the form

$$\widetilde{f}_{h,w}(x_0) = \sum_{x\in\mathbf{U}_{x_0,h}} w(x)Y(x), \tag{5}$$

where the unknown weights satisfy

$$w(x) \geq 0 \quad \text{and} \quad \sum_{x\in\mathbf{U}_{x_0,h}} w(x) = 1. \tag{6}$$

The usual bias plus variance decomposition of the Mean Square Error gives

$$\mathbb{E}\left(\widetilde{f}_{h,w}(x_0) - f(x_0)\right)^2 = Bias^2 + Var, \tag{7}$$

with

$$Bias^2 = \left(\sum_{x\in\mathbf{U}_{x_0,h}} w(x)\left(f(x) - f(x_0)\right)\right)^2$$

and

$$Var = \sigma^2 \sum_{x\in\mathbf{U}_{x_0,h}} w(x)^2.$$

The decomposition (7) is commonly used to construct asymptotically minimax estimators over some given classes of functions in the nonparametric function estimation. With our approach the bias term $Bias^2$ will be bounded in terms of the unknown function f itself. As a result we obtain some "oracle" weights w adapted to the unknown function f at hand, which will be estimated further using data patches of the image Y.

First, we address the problem of determining the "oracle" weights. With this aim denote

$$\rho_{f,x_0}(x) \equiv |f(x) - f(x_0)|. \tag{8}$$

Note that the value $\rho_{f,x_0}(x)$ characterizes the variation of the image brightness of the pixel x with respect to the pixel x_0. From the decomposition (7), we easily obtain a tight upper bound in terms of ρ_{f,x_0}:

$$\mathbb{E}\left(\widetilde{f}_h(x_0) - f(x_0)\right)^2 \leq g_{\rho_{f,x_0}}(w), \tag{9}$$

where

$$g_{\rho_{f,x_0}}(w) = \left(\sum_{x \in \mathbf{U}_{x_0,h}} w(x) \rho_{f,x_0}(x) \right)^2 + \sigma^2 \sum_{x \in \mathbf{U}_{x_0,h}} w(x)^2. \tag{10}$$

From the following theorem we can obtain the form of the weights w which minimize the function $g_{\rho_{f,x_0}}(w)$ under the constraints (6) in terms of $\rho_{f,x_0}(x)$. Introduce the strictly increasing function

$$M_{\rho_{f,x_0}}(t) = \sum_{x \in \mathbf{U}_{x_0,h}} \rho_{f,x_0}(x)(t - \rho_{f,x_0}(x))^+, \quad t \geq 0.$$

Let K_{tr} be the usual triangular kernel:

$$K_{\text{tr}}(t) = (1 - |t|)^+, \quad t \in \mathbb{R}^1. \tag{11}$$

Theorem 1. Assume that $\rho_{f,x_0}(x), x \in \mathbf{U}_{x_0,h}$, is a non-negative function. Then the unique weights which minimize $g_{\rho_{f,x_0}}(w)$ subject to (6) are given by

$$w_{\rho_{f,x_0}}(x) = \frac{K_{\text{tr}}(\frac{\rho_{f,x_0}(x)}{a})}{\sum\limits_{y \in \mathbf{U}_{x_0,h}} K_{\text{tr}}(\frac{\rho_{f,x_0}(x)}{a})}, \quad x \in \mathbf{U}_{x_0,h}, \tag{12}$$

where the bandwidth $a > 0$ is the unique solution in $(0, \infty)$ of the equation

$$M_{\rho_{f,x_0}}(a) = \sigma^2. \tag{13}$$

Remark 1. The value of $a > 0$ can be calculated as follows. We sort the set $\{\rho_{f,x_0}(x) \,|\, x \in \mathbf{U}_{x_0,h}\}$ in the ascending order $0 = \rho_1 \leq \rho_2 \leq \cdots \leq \rho_M < \rho_{M+1} = +\infty$, where $M = \text{Card}\, \mathbf{U}_{x_0,h}$. Let

$$a_k = \frac{\sigma^2 + \sum\limits_{i=1}^{k} \rho_i^2}{\sum\limits_{i=1}^{k} \rho_i}, \quad 1 \leq k \leq M, \tag{14}$$

and

$$\begin{aligned} k^* &= \max\{1 \leq k \leq M \,|\, a_k \geq \rho_k\} \\ &= \min\{1 \leq k \leq M \,|\, a_k < \rho_k\} - 1, \end{aligned} \tag{15}$$

with the convention that $a_k = \infty$ if $\rho_k = 0$ and that $\min \varnothing = M + 1$. Then the solution $a > 0$ of (13) can be expressed as $a = a_{k^*}$; moreover, k^* is the unique integer $k \in \{1, \cdots, M\}$ such that $a_k \geq \rho_k$ and $a_{k+1} < \rho_{k+1}$ if $k < M$.

Let $x_0 \in \mathbf{I}$. Using the optimal weights given by Theorem 1, we first introduce the following non computable approximation of the true image, called "oracle":

$$f_h^*(x_0) = \frac{\sum_{x \in \mathbf{U}_{x_0,h}} K_{\text{tr}}(\frac{\rho_{f,x_0}(x)}{a}) Y(x)}{\sum\limits_{y \in \mathbf{U}_{x_0,h}} K_{\text{tr}}(\frac{\rho_{f,x_0}(x)}{a})}, \tag{16}$$

where the bandwidth a is the solution of the equation $M_{\rho_{f,x_0}}(a) = \sigma^2$. A computable filter can be obtained by estimating the unknown function $\rho_{f,x_0}(x)$ and the bandwidth a from data pathes.

Let $h > 0$ and $\eta > 0$ be fixed numbers. For any $x_0 \in \mathbf{I}$ and any $x \in \mathbf{U}_{x_0,h}$ consider the distance between the data patches $\mathbf{Y}_{x,\eta} = (Y(y))_{y \in \mathbf{U}_{x,\eta}}$ and $\mathbf{Y}_{x_0,\eta} = (Y(y))_{y \in \mathbf{U}_{x_0,\eta}}$ defined by

$$d^2(\mathbf{Y}_{x,\eta}, \mathbf{Y}_{x_0,\eta}) = \frac{1}{m} \left\| \mathbf{Y}_{x,\eta} - \mathbf{Y}_{x_0,\eta} \right\|_2^2,$$

where $m = \text{card}\, \mathbf{U}_{x,\eta}$, and $\left\| \mathbf{Y}_{x,\eta} - \mathbf{Y}_{x_0,\eta} \right\|_2^2 = \sum\limits_{\|z\|_\infty \leq h} (Y(x+z) - Y(x_0+z))^2$ which measures the similarity between the data patches $\mathbf{Y}_{x,\eta}$ and $\mathbf{Y}_{x_0,\eta}$. Our simulations show that a convenient approximation of $\rho_{f,x_0}(x)$ is given by

$$\widehat{\rho}_{x_0}(x) = \left(d(\mathbf{Y}_{x,\eta}, \mathbf{Y}_{x_0,\eta}) - \sqrt{2}\sigma \right)^+. \tag{17}$$

A theoretical justification for this choice is given in a convergence theorem that is not presented here.

Thus our *Optimal Weights Filter* is defined by

$$\widehat{f}(x_0) = \widehat{f}_{h,\eta}(x_0) = \frac{\sum_{x \in \mathbf{U}_{x_0,h}} K_{\text{tr}}(\frac{\widehat{\rho}_{x_0}(x)}{\widehat{a}}) Y(x)}{\sum\limits_{y \in \mathbf{U}_{x_0,h}} K_{\text{tr}}(\frac{\widehat{\rho}_{x_0}(x)}{\widehat{a}})}, \tag{18}$$

where the bandwidth $\widehat{a} > 0$ is the solution of the equa-

Algorithm 1: Optimal weights filter.

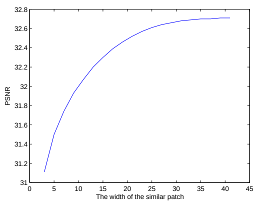
Repeat for each $x_0 \in \mathbf{I}$:
  give an initial value of $\widehat{a}$: $\widehat{a} = 1$ (it can be an arbitrary positive number).
  compute $\{\widehat{\rho}_{x_0}(x) \,|\, x \in \mathbf{U}_{x_0,h}\}$ by (17)
compute the bandwidth $\widehat{a}$ at x_0
  reorder $\{\widehat{\rho}_{x_0}(x) \,|\, x \in \mathbf{U}_{x_0,h}\}$ as increasing sequence, say
    $\widehat{\rho}_{x_0}(x_1) \leq \widehat{\rho}_{x_0}(x_2) \leq \cdots \leq \widehat{\rho}_{x_0}(x_M)$
  loop from $k = 1$ to M
    if $\sum_{i=1}^{k} \widehat{\rho}_{x_0}(x_i) > 0$
      if $\frac{\sigma^2 + \sum_{i=1}^{k} \widehat{\rho}_{x_0}^2(x_i)}{\sum_{i=1}^{k} \widehat{\rho}_{x_0}(x_i)} \geq \widehat{\rho}(x_k)$
        then $\widehat{a} = \frac{\sigma^2 + \sum_{i=1}^{k} \widehat{\rho}_{x_0}^2(x_i)}{\sum_{i=1}^{k} \widehat{\rho}_{x_0}(x_i)}$
    else quit loop
  else continue loop
  end loop
/compute the estimated weights $\widehat{w}$ at x_0
  compute $\widehat{w}(x_i) = \frac{K_{\text{tr}}(1 - \widehat{\rho}_{x_0}(x_i)/\widehat{a})^+}{\sum_{x_i \in \mathbf{U}_{x_0,h}} K_{\text{tr}}(1 - \widehat{\rho}_{x_0}(x_i)/\widehat{a})^+}$
/compute the filter $\widehat{f}$ at x_0
  compute $\widehat{f}(x_0) = \sum_{x_i \in \mathbf{U}_{x_0,h}} \widehat{w}(x_i) Y(x_i)$.

tion $M_{\widehat{\rho}_{x_0}}(\widehat{a}) = \sigma^2$, which can be calculated as in Remark 1 with $\rho_{f,x_0}(x)$ and a replaced by $\widehat{\rho}_{x_0}(x)$ and $\widehat{a}$ respectively. We end this section by giving an algorithm for computing the filter (18). The input values of the algorithm are the image $Y(x)$, $x \in \mathbf{I}$, the standard derivation σ of the Gaussian noise and two numbers m and M representing the sizes of data patches and search windows respectively (cf. (3) and (4)).

To avoid the undesirable border effects in simulations, we mirror the image outside the image limits, that is, we extend the image outside the image limits symmetrically with respect to the border. At the corners, the image is extended symmetrically with respect to the corner pixels.

The implementation of the proposed algorithm is straightforward. Notice that an important issue in the Non-Local Means Filter is the choice of the bandwidth parameter in the Gaussian kernel; our algorithm has the advantage that it automatically calculates the bandwidth.

A detailed analysis of the performance of our filter is given in Section 3 where the numerical simulations show that our filter outperforms the classical Non-Local Means Filter.



$\sigma = 20$

Figure 1: The evolution of PSNR value as a function of the size of data patches.

3 SIMULATIONS

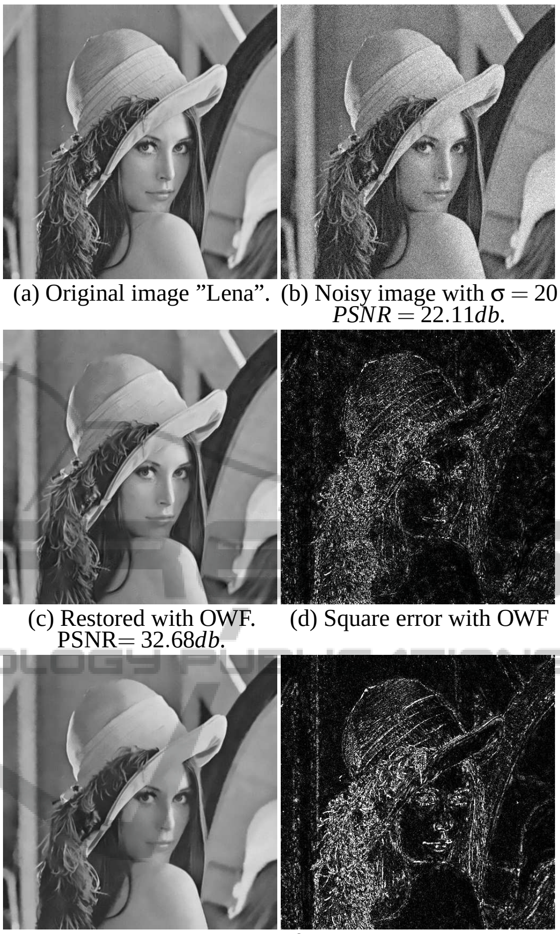
In this section we show the numerical performance of the Optimal Weights Filter by simulation results.

The performance of the Optimal Weights Filter $\widehat{f}_{h,\eta}(x_0)$ is measured by the usual Peak Signal-to-Noise Ratio (PSNR) in decibels (db) defined as

$$PSNR = 10\log_{10}\frac{255^2}{MSE},$$

$$MSE = \frac{1}{\operatorname{card}\mathbf{I}}\sum_{x\in\mathbf{I}}(f(x) - \widehat{f}_{h,\eta}(x))^2,$$

where f is the original image, and $\widehat{f}_{h,\eta}$ the estimated one.



(a) Original image "Lena". (b) Noisy image with $\sigma = 20$ $PSNR = 22.11db$.

(c) Restored with OWF. PSNR$= 32.68db$. (d) Square error with OWF

(e) Restored with NLMF. PSNR$= 31.51db$. (f) Square error with NLMF

Figure 2: Results of denoising "Lena" 512×512 image. Comparing (d) and (f) we see that the Optimal Weights Filter (OWF) captures more details than the Non-Local Means Filter (NLMF).

In the simulations, we sometimes use the smoothed version of the estimate of brightness variation $d_K(\mathbf{Y}_{x,\eta}, \mathbf{Y}_{x_0,\eta})$ instead of the non smoothed one $d(\mathbf{Y}_{x,\eta}, \mathbf{Y}_{x_0,\eta})$, defined by

$$d_K(\mathbf{Y}_{x,\eta}, \mathbf{Y}_{x_0,\eta}) = \frac{\|K(y)\cdot(\mathbf{Y}_{x,\eta} - \mathbf{Y}_{x_0,\eta})\|_2}{\sqrt{\sum_{y'\in\mathbf{U}_{x_0,\eta}}K(y')}},$$

where $K(y)$ are some weights defined on $\mathbf{U}_{x_0,\eta}$. The corresponding estimate of brightness variation $\rho_{f,x_0}(x)$ is given by

$$\widehat{\rho}_{K,x_0}(x) = \left(d_K(\mathbf{Y}_{x,\eta}, \mathbf{Y}_{x_0,\eta}) - \sqrt{2}\sigma\right)^+. \quad (19)$$

With the rectangular kernel

$$K_r(y) = \begin{cases} 1, & y \in \mathbf{U}_{x_0,\eta}, \\ 0, & \text{otherwise}, \end{cases} \quad (20)$$

Table 1: Performance of denoising algorithms when applied to test noisy (WGN) images.

	Images Sizes	Lena 512×512	Barbara 512×512	Boat 512×512	House 256×256	Peppers 256×256
σ	Method	PSNR	PSNR	PSNR	PSNR	PSNR
15	Our method($M = 13 \times 13$,$m = 27 \times 27$)	33.93db	32.31db	31.64db	34.09db	31.93db
	(Buades et al., 2005)	32.72db	31.67db	30.39db	33.82db	30.97db
	(Foi et al., 2004)	32.72db	29.61db	30.93db	33.18db	31.78db
	(Roth and Black, 2009)	33.29db	30.16db	31.27db	33.55db	32.06db
	(Hirakawa and Parks, 2006)	33.97db	32.55db	31.59db	33.82db	31.61db
	(Kervrann and Boulanger, 2008)	33.70db	31.80db	31.44db	34.08db	32.13db
	(Hammond and Simoncelli, 2008)	34.04db	32.25db	31.72db	33.72db	31.82db
	(Aharon et al., 2006)	33.71db	32.41db	31.77db	34.25db	32.20db
	(Dabov et al., 2007)	34.27db	33.00db	32.14db	34.94db	32.70db
20	Our method($M = 13 \times 13$,$m = 27 \times 27$)	32.68db	31.04db	30.30db	32.83db	30.61db
	(Buades et al., 2005)	31.51db	30.38db	29.32db	32.51db	29.73db
	(Foi et al., 2004)	31.43db	27.90db	39.61db	31.84db	30.30db
	(Roth and Black, 2009)	31.89db	28.28db	29.86db	32.29db	30.47db
	(Hirakawa and Parks, 2006)	32.69db	31.06db	30.25db	32.58db	30.21db
	(Kervrann and Boulanger, 2008)	32.64db	30.37db	30.12db	32.90db	30.59db
	(Hammond and Simoncelli, 2008)	32.81db	30.76db	30.41db	32.52db	30.40db
	(Aharon et al., 2006)	32.39db	30.84db	30.39db	33.10db	30.80db
	(Dabov et al., 2007)	33.05db	31.78db	30.88db	33.77db	31.29db
25	Our method($M = 13 \times 13$,$m = 27 \times 27$)	31.59db	29.92db	29.16db	31.95db	29.40db
	(Buades et al., 2005)	30.36db	29.19db	28.38db	31.16db	28.60db
	(Foi et al., 2004)	30.43db	26.62db	28.60db	30.75db	29.16db
	(Roth and Black, 2009)	30.57db	26.84db	28.57db	31.05db	29.17db
	(Hirakawa and Parks, 2006)	31.69db	29.89db	29.21db	31.60db	29.06db
	(Kervrann and Boulanger, 2008)	31.73db	29.24db	29.20db	32.22db	29.73db
	(Hammond and Simoncelli, 2008)	31.83db	29.58db	29.40db	31.54db	29.29db
	(Aharon et al., 2006)	31.36db	29.58db	29.32db	32.07db	29.67db
	(Dabov et al., 2007)	32.08db	30.72db	29.91db	32.86db	30.16db

we obtain exactly the distance $d\left(\mathbf{Y}_{x,\eta}, \mathbf{Y}_{x_0,\eta}\right)$ and the filter described in Section 2. Other smoothing kernels $K(y)$ used in the simulations are the Gaussian kernel

$$K_g(y) = \exp\left(-\frac{N^2 \|y - x_0\|_2^2}{2h_g}\right), \qquad (21)$$

where h_g is the bandwidth parameter, and the following kernel: for $y \in \mathbf{U}_{x_0,\eta}$,

$$K_0(y) = \sum_{k=\max(1,j)}^{N\eta} \frac{1}{(2k+1)^2} \qquad (22)$$

if $\|y - x_0\|_\infty = \frac{j}{N}$ for some $j \in \{0, 1, \cdots, N\eta\}$.

The best numerical results are obtained using $K(y) = K_0(y)$ in the definition of $\widehat{\rho}_{K,x_0}$. The values $m = 27 \times 27$ and $M = 13 \times 13$ are appropriate in most cases and a smaller data patch size m can be considered for processing piecewise smooth images. The comparison with several filters is given in Table 1. The PSNR values show that our approach is as good as more sophisticated methods, like (Hirakawa and Parks, 2006; Kervrann and Boulanger, 2008; Hammond and Simoncelli, 2008; Aharon et al., 2006), and is better than the filters proposed in (Foi et al., 2004; Roth and Black, 2009). Furthermore, our method is as simple as the Non-Local Means Filter and, with $K(y) = K_0(y)$, has only two parameters M and m which are the sizes of data patches and search windows. The proposed approach gives a denoising quality which is competitive with that of the recent method BM3D (Dabov et al., 2007).

The behavior of the PSNR in function of the size m of data patches is displayed in Figure 1 for "Lena" image. We fix $M = 13 \times 13$. For $\sigma = 20$, Figure 1 illustrates that the PSNR value increases as m varies between 3×3 and 41×41 (for which PSNR$= 32.71db$), and that it just changes slightly when m is sufficiently large (e.g. PSNR$= 32.68db$ when $m = 27 \times 27$). In our experimental results (cf. Table 1) we prefer $m = 27 \times 27$ as the choice $m = 41 \times 41$ is computationally expensive.

The potential of the estimation method is illustrated with the 512×512 image "Lena" (Figure 2(a)) corrupted by an additive white Gaussian noise (Figure 2(b), PSNR $= 22.10db$, $\sigma = 20$). We used the kernel $K_0(y)$ for computing the estimated brightness variation function $\widehat{\rho}_{K,x_0}$, which corresponds to the Optimal Weights Filter as defined in Section 2. In Figure 2(c), we can see that the noise is reduced in a natural manner and significant geometric features, fine textures, and original contrasts are visually well recovered with no undesirable artifacts (PSNR$= 32.68db$ for "Lena"). To better appreciate the accuracy of the restoration process, the square of the difference between the original image and the recovered image is shown in Figure 2(d), where the dark values correspond to a high-confidence estimate. As expected, pixels with a low level of confidence are located in the neighborhood of image discontinuities. For comparison, we show the image denoised by Non-Local Means Filter in Figures 2(e),(f). The overall visual impression and the numerical results are improved using our algorithm.

The Optimal Weights Filter seems to provide a feasible and rational method to detect automatically the details of images and take the proper weights for every possible geometric configuration of the image. The distribution of the weights inside the search window $\mathbf{U}_{x_0,h}$ depends on the estimated brightness variation function $\widehat{\rho}_{K,x_0}(x)$, $x \in \mathbf{U}_{x_0,h}$. If the estimated brightness variation $\widehat{\rho}_{K,x_0}(x)$ is less than $\widehat{a}$ (see Theorem 1), the similarity between patches is measured by a linear decreasing function of $\widehat{\rho}_{K,x_0}(x)$; otherwise it is zero. Thus $\widehat{a}$ acts as an automatic threshold.

4 CONCLUSIONS

We have proposed a new filter to remove Gaussian noise, based on optimization of weights in the weighted means approach. Our analysis shows that a triangular kernel is preferred rather then the Gaussian kernel. The proposed filter improves the usual Non-Local Means Filter both numerically and visually in denoising performance; it also has the advantage to be adaptive in the sense that it calculates automatically the good bandwidth of the triangular kernel (while in the Non-Local Means Filter the choice of the bandwidth parameter in the Gaussian kernel is delicate). We hope that the optimal weights that we deduced can also bring similar improvements for recently developed algorithms where the basic idea of the Non-Local means filter is used.

ACKNOWLEDGEMENTS

The authors are grateful to the reviewers for helpful comments and remarks. The work has been partially supported by the National Natural Science Foundation of China, Grant No. 11101039 and Grant No. 11171044.

REFERENCES

Aharon, M., Elad, M., and Bruckstein, A. (2006). *rmk-svd: An algorithm for designing overcomplete dictionaries for sparse representation. IEEE Trans. Signal Process.*, 54(11):4311–4322.

Buades, A., Coll, B., Morel, J., et al. (2005). A review of image denoising algorithms, with a new one. *SIAM Journal on Multiscale Modeling and Simulation*, 4(2):490–530.

Buades, T., Lou, Y., Morel, J., and Tang, Z. (2009). A note on multi-image denoising. In *Int. workshop on Local and Non-Local Approximation in Image Processing*, pages 1–15.

Dabov, K., Foi, A., Katkovnik, V., and Egiazarian, K. (2007). Image denoising by sparse 3-D transform-domain collaborative filtering. *IEEE Trans. Image Process.*, 16(8):2080–2095.

Donoho, D. and Johnstone, J. (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 81(3):425.

Foi, A., Katkovnik, V., Egiazarian, K., and Astola, J. (2004). A novel anisotropic local polynomial estimator based on directional multiscale optimizations . In *Proc. 6th IMA int. conf. math. in signal processing*, pages 79–82.

Hammond, D. and Simoncelli, E. (2008). Image modeling and denoising with orientation-adapted gaussian scale mixtures. *IEEE Trans. Image Process.*, 17(11):2089–2101.

Hirakawa, K. and Parks, T. (2006). Image denoising using total least squares. *IEEE Trans. Image Process.*, 15(9):2730–2742.

Katkovnik, V., Foi, A., Egiazarian, K., and Astola, J. (2010). From local kernel to nonlocal multiple-model image denoising. *Int. J. Comput. Vis.*, 86(1):1–32.

Kervrann, C. and Boulanger, J. (2008). Local adaptivity to variable smoothness for exemplar-based image regularization and representation. *Int. J. Comput. Vis.*, 79(1):45–69.

Lou, Y., Zhang, X., Osher, S., and Bertozzi, A. (2010). Image recovery via nonlocal operators. *J. Sci. Comput.*, 42(2):185–197.

Nazin, A., Roll, J., Ljung, L., and Grama, I. (2008). Direct weight optimization in statistical estimation and system identification. *System Identification and Control Problems (SICPRO08), Moscow*.

Polzehl, J. and Spokoiny, V. (2006). Propagation-separation approach for local likelihood estimation. *Probab. Theory Rel. Fields*, 135(3):335–362.

Roth, S. and Black, M. (2009). Fields of experts. *Int. J. Comput. Vision*, 82(2):205–229.

Sacks, J. and Ylvisaker, D. (1978). Linear estimation for approximately linear models. *Ann. Stat.*, 6(5):1122–1137.

Tomasi, C. and Manduchi, R. (1998). Bilateral filtering for gray and color images. In *Proc. Int. Conf. Computer Vision*, pages 839–846.

Yaroslavsky, L. P. (1985). Digital picture processing. An introduction. In *Springer-Verlag, Berlin*.