

Edge Preserving Image Denoising in Reproducing Kernel Hilbert Spaces

Pantelis Bouboulis, Sergios Theodoridis
*Department of Informatics and Telecommunications,
University of Athens,
Athens, Greece.*
{bouboulis,stheodor}@di.uoa.gr

Konstantinos Slavakis
*Department of Telecommunications
Science and Technology,
University of Peloponnese,
Tripolis, Greece.*
slavakis@uop.gr

Abstract

The goal of this paper is the development of a novel approach for the problem of Noise Removal, based on the theory of Reproducing Kernels Hilbert Spaces (RKHS). The problem is cast as an optimization task in a RKHS, by taking advantage of the celebrated semi-parametric Representer Theorem. Examples verify that in the presence of gaussian noise the proposed method performs relatively well compared to wavelet based technics and outperforms them significantly in the presence of impulse or mixed noise.

1 Introduction

The problem of noise removal from a digitized image is one of the most fundamental ones in digital image processing. So far, various techniques have been proposed to deal with it. Among the most important methodologies are, for example, the Wavelet-based image denoising methods, which dominates the research in recent years [2, 3]. In this paper we propose a novel approach which (to our knowledge) has not been considered before. We employ the well known powerful tool of kernels.

In kernel methodology the notion of the Reproducing Kernel Hilbert Space (RKHS) plays a crucial role. A RKHS, is a rich construct (roughly, a smooth space with an inner product), which has been proven to be a very powerful tool for non linear processing [9, 11]. In the denoising problem, we exploit a useful property of RKHS, the *representer theorem* [9]. It states that the minimizer of any optimization task in \mathcal{H} , with a cost function of a certain type, has a finite representation in \mathcal{H} . We recast the image denoising problem as an optimization task of this type and use the semi-parametric version of the representer theorem. The latter, allows

for explicit modeling of the edges in an image. In such a way we can deal with the smoothness which is, implicitly, imposed by the "smooth" nature of RKHS.

Though there has been some work exploring the use of kernels in the denoising problem, the methodology presented here is fundamentally different. In [10], the notion of kernel regression has been adopted. The original image is formulated as a Taylor approximation series around a center, \mathbf{x}_i , and data adaptive kernels are used, as weighted factors, to penalize distances away from \mathbf{x}_i . In a relatively similar context, kernels have been employed by other well known denoising methods (such as [1]). Kernels were also used in the context of RKHS in [6, 5]. However, the obtained results were not satisfying, especially around edges. It is exactly this drawback that is addressed by our method.

2 Mathematical Preliminaries

We start with some basic definitions regarding RKHS. Let X be a non empty set with $\mathbf{x}_1, \dots, \mathbf{x}_N \in X$. Consider a Hilbert space \mathcal{H} of real valued functions f defined on a set X , with a corresponding inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}}$. We will call \mathcal{H} as a *Reproducing Kernel Hilbert Space* - RKHS, if there exists a function, known as kernel, $\kappa : X \times X \rightarrow \mathbb{R}$ with the following two properties:

1. For every $\mathbf{x} \in X$, $\kappa(\mathbf{x}, \cdot)$ belongs to \mathcal{H} .
2. κ has the so called *reproducing property*, i.e. $f(\mathbf{x}) = \langle f, \kappa(\mathbf{x}, \cdot) \rangle_{\mathcal{H}}$, for all $f \in \mathcal{H}$. In particular $\kappa(\mathbf{x}, \mathbf{y}) = \langle \kappa(\mathbf{x}, \cdot), \kappa(\mathbf{y}, \cdot) \rangle_{\mathcal{H}}$.

It can be shown that the kernel κ produces the entire space \mathcal{H} , i.e. $\mathcal{H} = \text{span}\{\kappa(\mathbf{x}, \cdot) | \mathbf{x} \in X\}$. There are several kernels that are used in practice (see [9]). In this work, we focus on one of the most widely used, the

Gaussian Kernel:

$$\kappa(\mathbf{x}, \mathbf{y}) = \exp\left(-\frac{\|\mathbf{x} - \mathbf{y}\|^2}{2\sigma^2}\right), \sigma > 0,$$

due to some additional properties that it admits.

One of the many powerful tools in kernel theory is the application of the semi-parametric representer theorem to *regularized risk minimization* problems (see [9]):

Theorem 2.1. Denote by $\Omega_1, \Omega_2 : [0, \infty) \rightarrow \mathbb{R}$, two strictly monotonic increasing functions, by \mathcal{X} a set and by $c : (\mathcal{X} \times \mathbb{R}^2)^m \rightarrow \mathbb{R} \cup \{\infty\}$ an arbitrary loss function. Furthermore, consider a set of M real-valued functions $\{\psi_k\}_{k=1}^M : \mathcal{X} \rightarrow \mathbb{R}$, with the property that the $N \times M$ matrix $(\psi_p(\mathbf{x}_n))_{n,p}$ has rank M . Then any $\tilde{f} := f + h$, with $f \in \mathcal{H}$ and $h \in \mathfrak{H} = \text{span}\{\psi_k\}$, minimizing the regularized risk functional

$$c((\mathbf{x}_1, z_1, f(\mathbf{x}_1)), \dots, (\mathbf{x}_N, z_N, f(\mathbf{x}_N))) + \Omega_1(\|f\|_{\mathcal{H}}) + \Omega_2(\|h\|_{\mathfrak{H}})$$

admits a representation of the form

$$\tilde{f}(\mathbf{x}) = \sum_{n=1}^N \alpha_n \kappa(\mathbf{x}_n, \mathbf{x}) + \sum_{k=1}^M \beta_k \psi_k(\mathbf{x}). \quad (1)$$

Usually the regularization term $\Omega(f)$ takes the form $\Omega(f) = \frac{\lambda}{2} \|f\|_{\mathcal{H}}^2$. In the case of the RKHS produced by the gaussian Kernel we can prove that

$$\|f\|_{\mathcal{H}} = \int_{\mathcal{X}} \sum_n \frac{\sigma^{2n}}{n!2^n} (O^n f(\mathbf{x}))^2 d\mathbf{x}, \quad (2)$$

with $O^{2n} = \Delta^n$ and $O^{2n+1} = \nabla \Delta^n$, Δ being the Laplacian and ∇ the gradient operator (see [9]). Thus, we see that the regularization term "penalizes" the derivatives of the minimizer. This results to a very smooth solution of the regularized risk minimization problem.

Note that according to theorem 2.1 the model of a function has two parts, one lying in the smooth RKHS space and another part h which gives rise to the second term in the expansion (1). It is exactly this term that is exploited by our method in order to explicitly model edges.

3 Application to the denoising problem

Let f be the original image and \hat{f} the noisy one (we consider them as continuous functions). Also, let $f_{i,j}$ and $\hat{f}_{i,j}$ be the restrictions of f and \hat{f} on the $N \times N$ orthogonal region centered at the pixel (i, j) of each image accordingly (N is an odd number). Our task is to

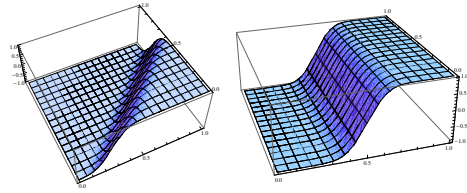


Figure 1. Two of the functions ψ_k that are used to represent edges.

find $f_{i,j}$ from the given samples of $\hat{f}_{i,j}$. For simplicity, we drop the i, j indices and consider $f_{i,j}$ and $\hat{f}_{i,j}$ (which from now on will be written as f and \hat{f}) as functions defined on $[0, 1]^2$ (and zero elsewhere). The pixel values of the digitized image are given by $f(x_n, y_m)$ and $\hat{f}(x_n, y_m)$ where $x_n = n/(N-1)$, $y_m = m/(N-1)$ for $n, m = 0, 1, \dots, N-1$.

We consider a set of real valued functions $\{\psi_k, k = 1, \dots, K\}$ with two variables suitable to represent edges; i.e., bivariate polynomials (which are controlled by the coefficients h_0, h_1, h_2, h_3) and functions of the form $\text{Erf}(a \cdot x + b \cdot y + c)$, where Erf is the error function,

$$\text{Erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt,$$

for several suitable choices of a, b and c (see figure 1). Thus we formulate the regularized risk minimization problem as follows:

$$\begin{aligned} & \underset{f \in \mathcal{H}, \beta \in \mathbb{R}^K, h \in \mathbb{R}^4}{\text{minimize}} \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} \left| f(x_n, y_m) + h_0 + h_1 x_n + \right. \\ & \left. + h_2 y_m + h_3 x_n y_m + \sum_{k=1}^K \beta_k \psi_k(x_n, y_m) - \hat{f}(x_n, y_m) \right| \\ & + \frac{\lambda}{2} \|f\|_{\mathcal{H}}^2 + \frac{\mu}{2} \sum_{k=1}^K |\beta_k|^2 + \frac{\mu_1}{2} \sum_{l=1}^3 h_l^2. \quad (3) \end{aligned}$$

Taking a closer look at the term $\frac{\lambda}{2} \|f\|_{\mathcal{H}}^2$ according to equation (2), one sees that we actually penalize the derivatives of f in a more influential fashion than the *total variation* scheme, which is often used in wavelet-based denoising and penalizes only the first order derivatives. It turns out that in our method the use of the L_1 norm in the cost function, in combination with regularization, results in sparse modeling with respect to the β coefficients. It should be noted that the use of the L_1 norm, also, in the regularization term leads to similar results.

The semi-parametric theorem 2.1 ensures that the

Table 1. Results on Boat corrupted by impulse noise.

Image	Noise	noisy PSNR	Kernel Denoising	BiShrink [2]	K-SVD [4]	SKR L_1 [10]	SKR [10]	BM3D [3]
Boat	20%	18.56 dB	32.36 dB	22.59 dB	26.46 dB	31.85 dB	28.35 dB	29.45 dB
	30%	16.77 dB	30.66 dB	25.07 dB	26.79 dB	30.85 dB	27.05 dB	28.29 dB
	40%	15.52 dB	29.14 dB	25.40 dB	26.08 dB	29.51 dB	25.85 dB	27.26 dB
	50%	14.55 dB	28.10 dB	25.09 dB	25.38 dB	27.73 dB	24.90 dB	26.61 dB

Table 2. Results on Lena corrupted by gaussian noise.

Image	Noise	noisy PSNR	Kernel Denoising	BiShrink [2]	BLS-GSM [8]	K-SVD [4]	SKR L_1 [10]	SKR [10]	BM3D [3]
Lena	$s = 10$	28.12 dB	33.98 dB	34.33 dB	35.60 dB	35.47 dB	32.66 dB	35.32 dB	35.93 dB
	$s = 20$	22.14 dB	31.12 dB	31.17 dB	32.65 dB	32.36 dB	29.23 dB	32.62 dB	33.00 dB
	$s = 30$	18.72 dB	29.11 dB	29.35 dB	30.50 dB	30.30 dB	26.60 dB	30.71 dB	31.21 dB

minimizer will have a finite representation of the form:

$$\tilde{f}(x, y) = \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} \alpha_{n,m} \kappa((x_n, y_m), (x, y)) + \sum_{k=1}^M \beta_k \psi_k(x, y) + h_0 + h_1 x + h_2 y + h_3 xy.$$

We can solve this problem using Polyak’s Projected Subgradient Method [7]. We fix the regularization parameter λ and adjust μ and μ_1 so that they take small values around edges and large values in smooth areas. In particular, as the algorithm moves from one pixel to the next, it decides whether the corresponding pixel centered region contains edges or not using the mean gradient of the specific region and then, it solves the corresponding minimization problem.

4 Experimental Results

Figure 2 and tables 1, 2 show the obtained results using our algorithm on the Lena and Boat (512×512) grayscale images. More experimental results, the code in C (for the proposed methodology), as well as details on the implementation may be found at <http://cgi.di.uoa.gr/~stheodor/kerden/index.htm>. The results were compared with those obtained using several state of the art wavelet-based denoising packages, which are available on the internet ([3, 10, 4, 1, 2]). The experiments show that the kernel approach performs equally well as the well-known BiShrink wavelet-based method [2] in the presence of Gaussian noise. However, it outperforms significantly the other denoising methods when impulse noise or mixed noise are considered (see figure 2). This enhanced performance is obtained at the cost of higher complexity, which is basically contributed by the optimization step, which is of the order of $O(MN)$ per pixel. Currently, more efficient optimization algorithms are considered. Moreover, the whole setting is open to a straightforward parallelization, when a parallel processing environment is available. This is also currently under consideration.

5 Conclusions

A novel denoising algorithm was presented based on the theory of RKHS. The semiparametric Representer Theorem was exploited in order to cope with the problems associated with the smoothing around edges, which is a common problem in almost all denoising algorithms. The comparative study against other denoising techniques, showed that significantly enhanced results are obtained in the case of impulse noise and mixed noise.

References

- [1] A. Buades, B. Coll, and J. M. Morrel. A review of image denoising algorithms, with a new one. *SIAM Multiscale Modelling and Simulation*, 4(2):490–530, 2005.
- [2] S. Cai and K. Li. Wavelet software at Brooklyn Poly. <http://taco.poly.edu/WaveletSoftware/index.html>.
- [3] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian. Image denoising by sparse 3d transform-domain collaborative filtering. *IEEE Tran. Im. Proc.*, 16(8):2080–2095, 2007.
- [4] M. Elad and M. Aharon. Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Tran. Im. Proc.*, 15(12):3736–3745, 2006.
- [5] K. Kim, M. O. Franz, and B. Scholkopf. Iterative kernel principal component analysis for image modeling. *IEEE Trans. Pattern Anal. Mach. Intell.*, 27(9):1351–1366, 2005.
- [6] D. Li. Support vector regression based image denoising. *Image Vision Comput.*, 27:623–627, 2009.
- [7] B. T. Polyak. *Introduction to Optimization*. New York: Optimization Software, 1987.
- [8] J. Portilla, V. Strela, M. Wainwright, and E. P. Simoncelli. Image denoising using scale mixtures of gaussians in the wavelet domain. *IEEE Trans. Im. Proc.*, 12(11):1338–1351, 2003.
- [9] B. Scholkopf and A. J. Smola. *Learning with Kernels*. MIT press, 2002.
- [10] H. Takeda, S. Farsiu, and P. Milanfar. Kernel regression for image processing and reconstruction. *IEEE Tran. Im. Proc.*, 16(2):349–366, 2007.
- [11] S. Theodoridis and K. Koutroumbas. *Pattern Recognition, 4th edition*. Academic Press, 2009.

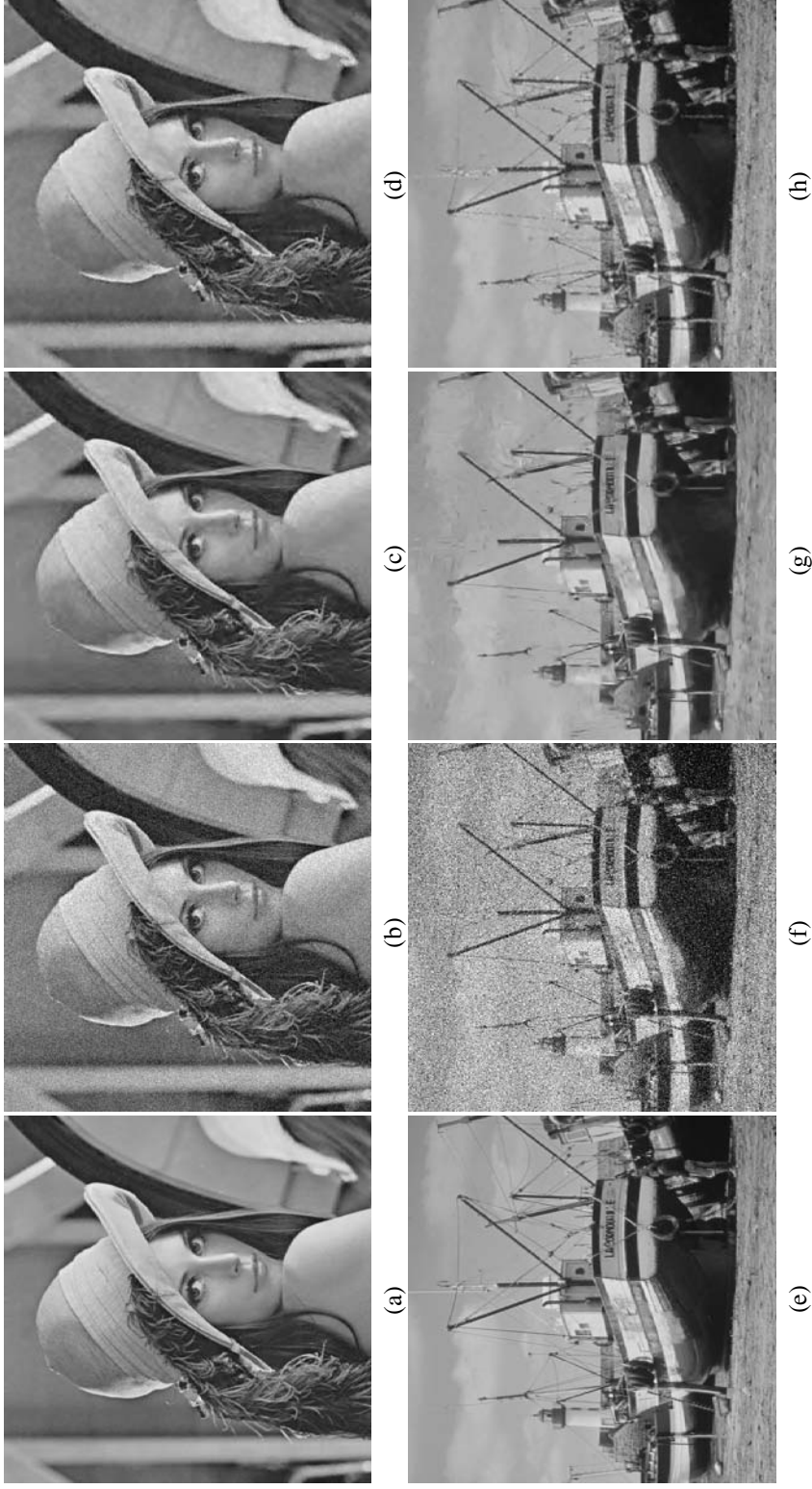


Figure 2. (a) Original Image, (b) Original with additive Gaussian Noise - PSNR=22.14 dB, (c) Wavelet BiShrink Denoising [2] - PSNR=31.17 dB, (d) Kernelized Denoising - PSNR=31.12 dB, (e) Original Image with additive Impulse Noise - PSNR=15.52 dB, (f) BM3D Denoising [3] - PSNR=27.26 dB, (g) Kernelized Denoising - PSNR=29.14 dB, (h) Wavelet BiShrink Denoising [2] - PSNR=22.14 dB.