ROBUST IMAGE DENOISING IN RKHS VIA ORTHOGONAL MATCHING PURSUIT

*Pantelis Bouboulis, George Papageorgiou, Sergios Theodoridis**

Department of Informatics and Telecommunications
University of Athens Athens, Greece, 157 84
Emails: panbouboulis@gmail.com, geopapag, stheodor@di.uoa.gr,

ABSTRACT

We present a robust method for the image denoising task based on kernel ridge regression and sparse modeling. Added noise is assumed to consist of two parts. One part is impulse noise assumed to be sparse (outliers), while the other part is bounded noise. The noisy image is divided into small regions of interest, whose pixels are regarded as points of a two-dimensional surface. A kernel based ridge regression method, whose parameters are selected adaptively, is employed to fit the data, whereas the outliers are detected via the use of the increasingly popular orthogonal matching pursuit (OMP) algorithm. To this end, a new variant of the OMP rationale is employed that has the additional advantage to automatically terminate, when all outliers have been selected.

Index Terms— image denoising, OMP, OMP termination criteria, kernels, Reproducing Kernel Hilbert Space, outliers, Kernel Ridge Regression

1. INTRODUCTION

The problem of image denoising is one of the most fundamental ones in the area of image processing. Many techniques have been proposed to deal with it, ranging from the popular wavelet-based image denoising framework (e.g., [1, 2, 3]), the maximum likelihood estimation methods (e.g., [4]) and the methods based on Partial Differential Equations (e.g., [5]), to specific methods for impulse detection (e.g., [6]) and methods of non linear modeling using regression and/or local expansion approximation techniques [7].

Another path, that has been considered in the respective literature, is the adoption of the rich and powerful machine learning toolbox. Following this rationale, in [8] and [9], a support vector regression (SVR) approach is considered to fit the image pixels, while in [10] the kernel principal components of an image are extracted and the respective expansion is truncated to produce the denoising effect. On the other hand, in [11, 12], a kernel-based regularized regression task is considered, exploiting the semi-parametric representer theorem as a means to model and preserve the sharp edges of the image, following a function expansion rationale as in K-SVD algorithm [13].

In the present paper, we adopt an alternative strategy. We assume that the noisy image is modeled as a sum of three parts: a) a function, f, lying in a *Reproducing Kernel Hilbert Space* (RKHS), which represents the noise-free picture, b) a bounded noise vector, $\boldsymbol{\eta}$, and c) a sparse vector, $\boldsymbol{u}$, which represents the impulses that are randomly distributed over the image samples. An iterative procedure is employed, that cycles over a regularized risk minimization task, which identifies the smooth part of the expansion (i.e., f) and extracts the bounded noise $\boldsymbol{\eta}$, and an *Orthogonal Matching Pursuit* (OMP) selection technique (e.g., [14, 15]) that approximates the sparse vector $\boldsymbol{u}$. Besides the sparse vector approximation via the OMP, that targets the outliers, we also consider a nonlinear modeling in a RKHS, which captures the fine details of the image (including the sharp edges), and a regularization mechanism to guard against overfitting (i.e., considering the outliers as some sort of sharp edges). The proposed scheme is based on the recently introduced *Kernel Regularized OMP* (KROMP) [16]. However, in this paper we also adopt an automatic termination criterion for the OMP mechanism. This specific variant of the OMP is able to automatically decide whether all outliers have been detected or not and hence terminate the selection procedure without prior knowledge of the bound of $\boldsymbol{\eta}$. It is demonstrated that the proposed method offers comparable performance to [11], albeit at significantly reduced complexity. In the case of mixed noise comprised of both Gaussian and impulse components, the present method considerably outperforms other state of the art wavelet based techniques. Moreover, in this paper, theoretical results regarding several important properties of KROMP are established, for the first time. These include conditions for the exact recovery of the support of the sparse vector, $\boldsymbol{u}$, as well as bounds for the final solution of KROMP.

*This research has been co-financed by the European Union (European Social Fund ESF) and Greek national funds through the Operational Program Education and Lifelong Learning of the National Strategic Reference Framework (NSRF) - Research Funding Program: Aristeia I: 621.

978-1-4799-3696-0/14/$31.00 ©2014 IEEE

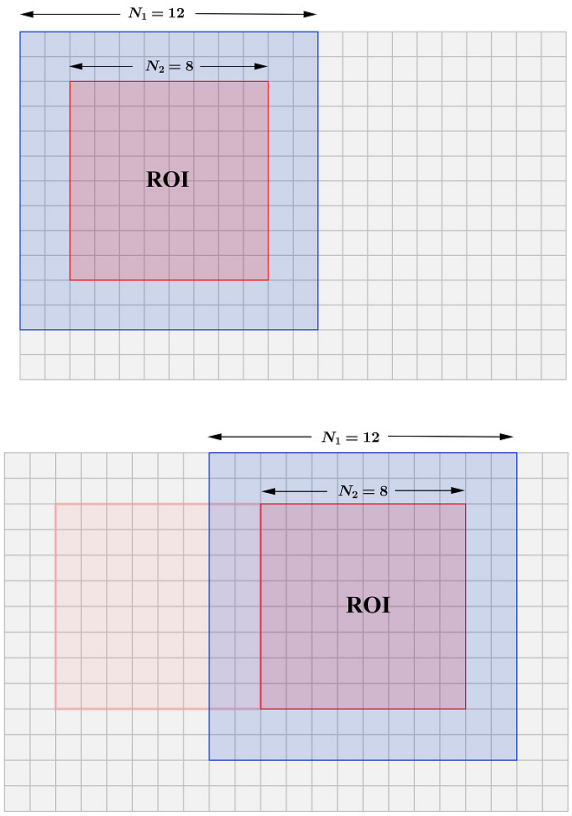

Fig. 1. Although each ROI contains $N_1 \times N_1$ pixels (blue), only $N_2 \times N_2$ of them (red) are used for the reconstruction of the image.

2. PROBLEM MODELING

Suppose we are given a set of data of the form:

$$\mathcal{D} = \{(k, l, p_{k,l}),\ k, l = 0, 1, \ldots, N-1\},$$

such that $p_{i,j} = 0, 1, \ldots, 255$, representing a $N \times N$ *region of interest* (ROI) of a specific noisy grayscale picture. Rearranging the data points, we may equivalently assume that $\mathcal{D}$ takes the form $\mathcal{R} = \{(\boldsymbol{x}_i, y_i), i = 1, \ldots, N^2\}$, where each $\boldsymbol{x}_i$, y_i are equal to $\boldsymbol{x}_i = (k/(N-1), l/(N-1))$ and $y_i = p_{k,l}$, respectively, for some $k, l \in \{0, 1, \ldots, N-1\}$. Moreover, assume that

$$y_i = f(\boldsymbol{x}_i) + \eta_i + u_i, \tag{1}$$

for $i = 1, \ldots N^2$, where f is a nonlinear function representing the noisy-free image, $\boldsymbol{\eta}$ is an unobservable bounded noise sequence and $\boldsymbol{u}$ a sparse vector of outliers.

In this paper we consider the case, where the nonlinear function, f, lies in a *Reproducing Kernel Hilbert Space* (RKHS) induced by the Gaussian kernel, $\kappa(\boldsymbol{x}, \boldsymbol{y}) = \exp(-\|\boldsymbol{x} - \boldsymbol{y}\|^2/\sigma^2)$. RKHS, [17, 18], are inner product function spaces with the following reproducing property: $f(\boldsymbol{x}) = \langle f, \kappa(\cdot, \boldsymbol{x})\rangle_{\mathcal{H}}$. These spaces have been proved to be a very powerful tool in statistical learning [18, 19, 20].

In contrast to our method, prior works, e.g., [11, 8, 9, 10], have considered modeling the noisy-free image in a RKHS, as

$$y_i = f(\boldsymbol{x}_i) + \eta_i, \tag{2}$$

without explicit modeling of the outliers and minimizing for different kinds of empirical loss. In this context, these techniques are only suitable for removing white Gaussian noise, as impulse noise, or any other type of noise that follows a heavy tailed distribution, usually causes the model to overfit, if the model parameters are chosen so that the fine details of the image are preserved. On the other hand, if the model parameters are chosen so that the smoothing effect is enhanced to remove the complete set of outliers, the fine details are lost.

To resolve these issues, the present work considers a more robust approach. The outliers are explicitly modeled as a sparse vector, $\boldsymbol{u}$, and the popular orthogonal matching pursuit rationale is adopted to identify its support. Inspired by the celebrated representer theorem [21], we assume that f takes the form: $f(\boldsymbol{x}) = \sum_{j=1}^{N^2} a_j \kappa(\cdot, \boldsymbol{x}_j) + c$, and solve the following optimization task:

$$\begin{aligned} \underset{\boldsymbol{a}, \boldsymbol{u} \in \mathbb{R}^{N^2}, c \in \mathbb{R}}{\text{minimize}} \quad & \|\boldsymbol{u}\|_0 \\ \text{subject to} \quad & \sum_{i=1}^{N^2} \left(y_i - \sum_{j=1}^{N^2} a_j \kappa(\boldsymbol{x}_i, \boldsymbol{x}_j) - c - u_j \right)^2 \\ & + \lambda(c^2 + \|\boldsymbol{a}\|^2) \leq \epsilon, \end{aligned} \tag{3}$$

where $\lambda > 0$ is the regularization parameter. Using matrix notation, (3) can be shown to take the form:

$$\begin{aligned} \underset{\boldsymbol{a}, \boldsymbol{u} \in \mathbb{R}^{N^2}, c \in \mathbb{R}}{\text{minimize}} \quad & \|\boldsymbol{u}\|_0 \\ \text{subject to} \quad & \|\boldsymbol{y} - A\boldsymbol{z}\|^2 + \lambda \boldsymbol{z}^T B \boldsymbol{z} \leq \epsilon, \end{aligned} \tag{4}$$

where

$$A = \begin{bmatrix} K & \mathbf{1} & I_n \end{bmatrix},\ \boldsymbol{z} = \begin{bmatrix} \boldsymbol{\alpha} \\ c \\ \boldsymbol{u} \end{bmatrix},\ B = \begin{bmatrix} I_{N^2} & \mathbf{0} & O_{N^2} \\ \mathbf{0}^T & 1 & \mathbf{0}^T \\ O_{N^2} & \mathbf{0} & O_{N^2} \end{bmatrix},$$

while I_{N^2} denotes the unitary matrix, $\mathbf{1}$ and $\mathbf{0}$ the vectors of ones and zeros respectively and O_{N^2} the all zero square matrix. In other words, our goal is to find the sparsest vector of outliers, $\boldsymbol{u}$, such that the mean square error between the actual noisy data and the estimated noise-free data plus the outliers remains below a certain tolerance, ϵ, while, at the same time, the solution remains "simple" (i.e., smooth). The latter is ensured by the regularization part of (4). The solution of (4), which is computed by algorithm 1, defines the noisy-free image, i.e., $\boldsymbol{y}_{est} = K\boldsymbol{a} + c\mathbf{1}$. Note that although (4) employs the ℓ_0 pseudo-norm, it has been shown that it can be efficiently solved using greedy techniques, like the popular OMP (another solution for (4), employing a LASSO-like rationale can

be found in [22]). The following theorems (their proofs are omitted due to lack of space) establish that algorithm 1 converges and that, under certain reasonable assumptions, it is able to recover the "correct" support of $\boldsymbol{u}$ with bounded reconstruction error.

Theorem 1 *The norm of the residual vector $r^{(k)} = A_{ac}\boldsymbol{z}^{(k)} - \boldsymbol{y}$ in algorithm 1 is strictly decreasing. Moreover, the algorithm will always converge.*

Theorem 2 *Assuming that $\boldsymbol{y}$ admits a representation of the form $\boldsymbol{y} = K\boldsymbol{a}_0 + c_0\mathbf{1} + \boldsymbol{u}_0 + \eta$, where $\|\boldsymbol{u}\|_0 = S$ and $\|\eta\|_2 \leq \epsilon$, then algorithm 1 recovers the exact sparsity pattern of $\boldsymbol{u}$, after N steps, if*

$$\epsilon \leq \frac{\left| \min\limits_{k=1,2,\ldots,N^2} \{u_k, u_k \neq 0\} \right|}{2} - \|\boldsymbol{f}_0 - \boldsymbol{f}^{(0)}\|_2,$$

with $\boldsymbol{f}_0 = \begin{pmatrix} K & \mathbf{1} \end{pmatrix} \cdot \begin{pmatrix} \boldsymbol{a}_0 \\ c_0 \end{pmatrix}$ *and* $\boldsymbol{f}^{(0)} = \begin{pmatrix} K & \mathbf{1} \end{pmatrix} \cdot \begin{pmatrix} \boldsymbol{a}^{(0)} \\ c^{(0)} \end{pmatrix}$.

Theorem 3 *Assuming that $\boldsymbol{y}$ admits a representation as in theorem 2, then the approximation error of algorithm 1, after S steps, is bounded by*

$$\|\boldsymbol{z}^{(N)} - \boldsymbol{z}_0\| \leq \frac{|\lambda| \|(\boldsymbol{a}_0 \quad c_0)\| + \sqrt{N}\left(1 + \|(K \quad \mathbf{1})\|_2\right)\epsilon}{\lambda_{min}\left(\Omega^T\Omega + \lambda I_0\right)},$$

where $\lambda_{min}\left(\Omega^T\Omega + \lambda I_0\right)$ is the minimum eigenvalue of the matrix $\Omega^T\Omega + \lambda I_0$, where

$$\Omega = \begin{pmatrix} K & \mathbf{1} & I \end{pmatrix}, \quad I_0 = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & I \end{pmatrix}.$$

The user-selected parameters ϵ, λ, control the quality of the reconstruction regarding the outliers and the bounded noise, respectively. Small values of ϵ model all noise samples (even those originating from a Gaussian source) as impulses, filling up the vector $\boldsymbol{u}$, which will no longer be sparse. Similarly, if λ is very small, f will closely fit the noisy data including some possible outliers (overfitting). On the other hand, if ϵ is very large, only a handful of outliers (possibly none) will be detected. If λ is large, then f will be smooth resulting to a blurry picture. In the following, we present a method that automatically computes the best ϵ thus only the choice λ is left to the user.

Finally, we have to point out that the accuracy of fit of any kernel-based regression technique drops near the borders of the data set. The proposed algorithmic scheme takes this important fact into consideration, removing these points from the reconstructed image.

Algorithm 1 : Automated Kernel Regularized OMP (autoKROMP)

Require: K, $\boldsymbol{y}$, λ, ϵ
  $n = N^2$.
  Initialization: $k := 0$
  $S_{ac} = \{1, 2, ..., n+1\}$, $S_{inac} = \{n+2, ..., 2n+1\}$
  $A_{ac} = [K\ \mathbf{1}]$, $A_{inac} = I_n = [\boldsymbol{e}_1\ \cdots\ \boldsymbol{e}_n]$
  Solve: $\boldsymbol{z}^{(0)} := \arg\min_{\boldsymbol{z}} ||A_{ac}\boldsymbol{z} - \boldsymbol{y}||_2^2 + \lambda||\boldsymbol{\alpha}||_2^2 + \lambda c^2$
  Initial Residual: $\boldsymbol{r}^{(0)} = A_{ac}\boldsymbol{z}^{(0)} - \boldsymbol{y}$
  Construct a histogram of the values of the residual $|\boldsymbol{r}^{(0)}|$, using $ceil(n/10)$ categories (intervals) of the same size.
  Set T_0 as the center of the first interval with the lowest frequency.
  while $\max\{|r_i^{(k)}|, i = 1, \ldots, n\} > T_0$ **do**
    $k := k + 1$
    Find: $j_k := \arg\max_{j \in S_{inac}} |r_j^{(k-1)}|$
    Update Support:
    $S_{ac} = S_{ac} \cup \{j_k\}$, $S_{inac} = S_{inac} - \{j_k\}$
    $A_{ac} = [A_{ac}\ \boldsymbol{e}_{j_k}]$
    Update Current solution:
    $\boldsymbol{z}^{(k)} := \arg\min_{\boldsymbol{z}} ||A_{ac}\boldsymbol{z} - \boldsymbol{y}||_2^2 + \lambda||\boldsymbol{\alpha}||_2^2 + \lambda c^2$
    Update Residual: $\boldsymbol{r}^{(k)} = A_{ac}\boldsymbol{z}^{(k)} - \boldsymbol{y}$
    Update T_0.
  end while

3. THE ALGORITHM

The proposed denoising procedure splits the noise image $\mathcal{P}$ into small square ROIs, $\mathcal{R}$, with dimensions $N_1 \times N_1$. Then, it applies the autoKROMP algorithm (see algorithm 1) sequentially to each $\mathcal{R}$ to find the noise free ROI $\tilde{\mathcal{R}}$. However, only the centered $N_2 \times N_2$ pixels are used to reconstruct the noise free image. As kernel based regression algorithms suffer from reduced accuracy at the points lying close to the borders of the respective regions, these are removed from the final result. Hence, in order to avoid gaps in the final reconstructed image, the ROIs contain overlapping sections (figure 1). The procedure requires two user defined parameters: (a) the regularization parameter, λ, that appears in (4) and (b) the Gaussian kernel parameter σ. The λ parameter regulates the smoothness of f, i.e., small values of λ increase the derivatives of f (e.g., more noise is included), while larger values lead to flat, less accurate solutions. For this reason the values of λ are chosen in an adaptive manner, so that smaller values of λ are used in areas of the picture that contain edges. While the user defines a single value for λ, say λ_0, the denoising scheme computes the mean gradient magnitude of each ROI and classifies them into three categories. The ROIs of the first class, i.e., those with the largest values of gradient magnitude, solve (4) using $\lambda = \lambda_0$, while for the other two classes the algorithm employs $\lambda = 5 * \lambda_0$ and $\lambda = 15 * \lambda_0$ respectively. Such choices came out after extensive experimentation, and the algorithm is not sensitive to them.

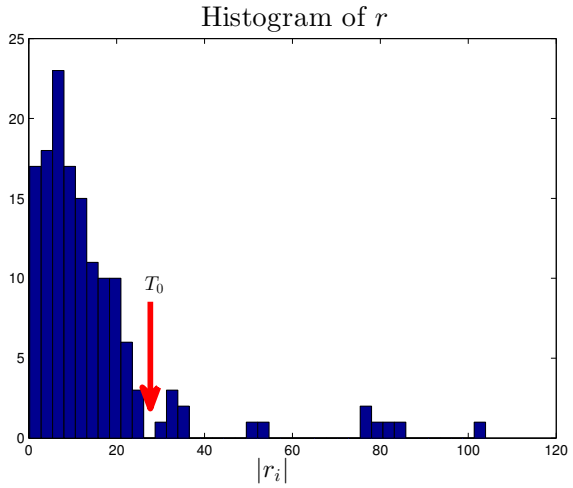

Fig. 2. Computation of T_0 for the stoping rule.

The autoKROMP algorithm cycles through an outlier selection criterion, that locates the position, say j_k, of the outlier most further away from the current estimation; then a regularized least squares minimization task estimates the vector $\boldsymbol{z}$ to be used at the next step i.e., $\boldsymbol{z}^{(n)} = \arg\min \|\boldsymbol{y} - A_{ac}\boldsymbol{z}\|^2 + \lambda \boldsymbol{z}^T B_{ac}\boldsymbol{z}$. The matrix A_{ac} is initialized at $[K\ \mathbf{1}]$ and augmented at each cycle by the vector e_{j_k}. Although the algorithm presented above solves one minimization task per iteration step, this can be bypassed, by applying a matrix decomposition on A_{ac} and employ Cholesky factorization arguments, that significantly simplify complexity. This is a common technique in OMP related methods and it is made possible as large blocks of the matrices A_{ac} and B_{ac} remain unchanged [16].

The presented algorithm exploits a very simple, yet effective automatic stopping rule. The main idea of this scheme is presented in figure 2. At each step, i, the algorithm constructs a histogram of the values of the residual, i.e., $|\boldsymbol{r}^{(i-1)}|$, and then computes the first interval with the lowest frequency. The center of that interval is denoted as T_i and the algorithm assumes that all residual's coordinates larger than $T_0^{(i)} = \min\{T_k, k = 1, \ldots, i\}$ correspond to outliers. The cycle stops, when the absolute values of all coordinates of the residual vector drop below $T_0^{(i)}$.

4. EXPERIMENTS

In this section, we demonstrate the effectiveness of the proposed algorithm in the denoising problem (figure 3). We compare autoKROMP with another kernel-based denoising method ([11]) and the popular BM3D wavelet denoising method [3]. The experiments have been carried out using the Lena and Boat images from the Waterloo Image Repository, which have been corrupted by synthetic noise comprising of a Gaussian part and randomly placed outliers. Table 1 shows that autoKROMP has comparable performance with the kernel-based method of [11], while at the same time it requires significantly less computational resources. In our experiments autoKROMP required only a few seconds to complete the task, which is comparable with that required by BM3D. Note that this is less than 1/20th of the time that the method of [11] needs for the same reason (recall that [11] employs a dictionary to model the edges of the image and solves the respective minimization problem using Polyak's Projected Subgradient Method). Compared to the BM3D method, autoKROMP is able to efficiently remove the majority of outliers, achieving higher PSNRs in most cases (e.g., the Lena and Boat images in table 1). Note that in all the presented experiments the parameters of the specific algorithms have been carefully tuned to get the best possible PSNRs. Specifically, the autoKROMP algorithm has been implemented using the following parameters: (a) for the three cases of the Lena image (see table 1a): $\sigma = 0.3, 0.3$ and 0.35, $\lambda = 1$, for each case respectively, and (b) for the boat image (table 1b): $\sigma = 0.25, 0.25$ and 0.3, $\lambda = 1$. In all cases the dimension for the ROIs were set equal to $N1 = 12$ and $N2 = 8$. The respective code (implemented in MatLab) can be found at bouboulis.mysch.gr/kernels.html.

Noise	autoKROMP	Kernel [11]	BM3D [3]
20 dB Gaussian + 5% outliers	32.05dB	32.34dB	29.70dB
20 dB Gaussian + 10% outliers	31.36dB	31.69dB	28.30dB
20 dB Gaussian + 15% outliers	30.48dB	30.55dB	26.30dB

(a)

Noise	autoKROMP	Kernel [11]	BM3D [3]
20 dB Gaussian + 5% outliers	29.61dB	29.85dB	27.82dB
20 dB Gaussian + 10% outliers	28.91dB	28.95dB	26.59dB
20 dB Gaussian + 15% outliers	28.28dB	28.37dB	25.00dB

(b)

Table 1. PSNRs of the reconstructed images using various denoising techniques for: (a) the Lena image and (b) the Boat image.

5. CONCLUSIONS

We presented a robust denoising approach based on Kernel Ridge Regression and sparse modeling via the OMP rationale. Our method cycles using an outlier identification step and a least squares minimization task that is solved iteratively using Cholesky decomposition. The algorithm terminates automatically using an effective termination criterion for the OMP mechanism, which can be used in any other OMP realization. The performance of the proposed denoising method was demonstrated in the case of mixed noise (consisting of both Gaussian and impulse components). Experiments showed that the present method can significantly outperform state of the art wavelet based methods with comparable complexity.

(a)

(b)

(c)

(d)

Fig. 3. The Lena image: (a) corrupted by Gaussian noise plus 5% impulses and reconstructed using (b) the BM3D method of [3] (PSNR = 29.7 dB), (c) autoKROMP (PSNR = 32.05 dB) and (d) the kernel based method of [11] (PSNR = 32.34 dB).

(a)

(b)

Fig. 4. The Boat image: (a) corrupted by Gaussian noise plus 5% impulses and (b) reconstructed using autoKROMP (PSNR = 29.61 dB).

(a)

(b)

Fig. 5. The Boat image: (a) corrupted by Gaussian noise plus 10% impulses and (b) reconstructed using autoKROMP (PSNR = 28.91 dB).

6. REFERENCES

[1] 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.*, vol. 12, no. 11, pp. 1338–1351, 2003.

[2] L. Sendur and I.W. Selesnick, "Bivariate shrinkage functions for wavelet-based denoising exploiting interscale dependency," *IEEE Trans. Signal Process.*, vol. 50, no. 11, pp. 2744–2756, 2002.

[3] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3d transform-domain collaborative filtering," *IEEE Tran. Im. Proc.*, vol. 16, no. 8, pp. 2080–2095, 2007.

[4] Jun Liu, Xue-Cheng Tai, Haiyang Huang, and Zhongdan Huan, "A weighted dictionary learning model for denoising images corrupted by mixed noise," *Image Processing, IEEE Transactions on*, vol. 22, no. 3, pp. 1108–1120, 2013.

[5] K. Seongjai, "PDE-based image restoration : A hybrid model and color image denoising," *IEEE Trans. Im. Proc.*, vol. 15, no. 5, pp. 1163–1170, 2006.

[6] R.Garnett, T. Huegerich, and C Chui, "A universal noise removal algorithm with an impulse detector," *IEEE Trans. Im. Proc.*, vol. 14, no. 11, pp. 1747–1754, 2005.

[7] H. Takeda, S. Farsiu, and P. Milanfar, "Kernel regression for image processing and reconstruction," *IEEE Tran. Im. Proc.*, vol. 16, no. 2, pp. 349–366, 2007.

[8] Dalong Li, "Support vector regression based image denoising," *Image Vision Comput.*, vol. 27, pp. 623–627, 2009.

[9] S. Zhang and Y. Chen, "Image denoising based on wavelet support vector machine," *IICCIAS 2006*.

[10] K. Kim, M. O. Franz, and B. Scholkopf, "Iterative kernel principal component analysis for image modeling," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 9, pp. 1351–1366, 2005.

[11] P. Bouboulis, K. Slavakis, and S. Theodoridis, "Adaptive kernel-based image denoising employing semi-parametric regularization," *IEEE Transactions on Image Processing*, vol. 19, no. 6, pp. 1465–1479, 2010.

[12] P. Bouboulis, K. Slavakis, and S. Theodoridis, "Edge preserving image denoising in reproducing kernel hilbert spaces," *Proceedings - International Conference on Pattern Recognition*, , no. 5596011, pp. 2660–2663, 2010.

[13] M. Elad and M. Aharon, "Image denoising via sparse and redundunt representations over learned dictionaries," *IEEE Tran. Im. Proc.*, vol. 15, no. 12, pp. 3736–3745, 2006.

[14] Alfred M. Bruckstein, David L. Donoho, and Michael Elad, "From sparse solutions of systems of equations to sparse modeling of signals and images," *SIAM review*, vol. 51, no. 1, pp. 34–81, 2009.

[15] Sergios Theodoridis, Yiannis Kopsinis, and Kostantinos Slavakis, "Sparsity-aware learning and compressed sensing: An overview," *http://arxiv.org/abs/1211.5231*.

[16] G. Papageorgiou, P. Bouboulis, and S. Theodoridis, "Robust kernel-based regression using orthogonal matching pursuit," in *MLSP*, September 2013.

[17] K. Slavakis, P. Bouboulis, and S. Theodoridis, "Online learning in reproducing kernel Hilbert spaces," *Elsevier's E-Reference Signal Processing*, vol. 1, section 5, article number 33, pp. 1–76, 2013 (to appear).

[18] B. Scholkopf and A.J. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization and Beyond*, MIT Press, 2002.

[19] J. Shawe-Taylor and N. Cristianini, *Kernel Methods for Pattern Analysis*, Cambridge University Press, 2004.

[20] S. Theodoridis and K. Koutroumbas, *Pattern Recognition*, Academic Press, 4^{th} edition, Nov. 2008.

[21] G. S. Kimeldorf and G. Wahba, "Some results on Tchebycheffian spline functions," *J. Math. Anal. Applic.*, vol. 33, pp. 82–95, 1971.

[22] Gonzalo Mateos and Georgios B. Giannakis, "Robust nonparametric regression via sparsity control with application to load curve data cleansing," *Signal Processing, IEEE Transactions on*, vol. 60, no. 4, pp. 1571–1584, 2012.