SIGNAL AND IMAGE DENOISING VIA SCALE-SPACE ATOMS

V. Bruni, B. Piccoli, D. Vitulano

Istituto per le Applicazione del Calcolo "M. Picone" - C.N.R.
Viale del Policlinico 137, 00161 Rome, Italy
phone: + (39) 06 88470224, fax: + (39) 06 4404306, email: bruni,piccoli, vitulano@iac.rm.cnr.it

ABSTRACT

In this paper a scale-space model for signal and image denoising is presented. The wavelet coefficients are split into overlapping atoms and their evolution law through scales is formally derived. This formalism accounts for both inter and intra-scale dependencies of wavelet coefficients and then it can be exploited for denoising. Experimental results show that the proposed algorithm outperforms the most effective wavelet based denoising techniques.

1. INTRODUCTION

The objective of a de-noising tool consists of recovering a signal f from its noisy version g, where

$$g(t) = f(t) + \upsilon(t), \quad t \in \mathbf{R}$$

and υ is an additive zero-mean Gaussian noise with variance σ^2. It is commonly solved by deriving a suitable representation of the signal in which discarding noise with simple operations, like thresholding. In fact, non linear approximation theory proves that the optimal expansion basis for a signal yields coefficients with a fast decay with respect to L^p norm, for some $p < 2$ [9]. This is one of the main reasons of the success of wavelets in signal representation. Nonetheless, the decay of the coefficients does not completely characterize a signal. Dependencies between adjacent coefficients are very important (for example edges in 2D signals), expecially in the noisy case. In fact, gaussian noise does not reveal any spatial correlation. For that reason new bases have been defined. They are oriented to capture some prefixed structures of the analysed signal, such as singularities, smooth contour curves, geometric flow, and so on [5, 8, 12]. However, intrascale dependencies are again lost if a simple thresholding is used for removing noise, yielding annoying artifacts in the recovered signal.

A step forward is represented by the characterization of singularity points in the wavelet scale-space domain [7, 9]. In fact, the location and the degree of singularities of the analyzed function can be derived by the decay of the corresponding wavelet coefficients using derivatives of a gaussian kernel (Chap. 6 of [9]). On the other hand, noise can be discriminated since it has an opposite behaviour through scales. For piecewise polynomial functions scale relationships can be emphasized using general kernels [7]. Nonetheless, both approaches suffer from the fact that singularities can be discriminated and then recovered by imposing a constraint on their minimum distance.

The aim of this paper is to provide a model for representing wavelet coefficients both in time and scale. It consists of splitting the wavelet transform of a signal into *overlapping atoms* that correspond to basic singularity points in the original signal, as shown in Fig. 1. The atoms are defined both in time and scale and can be marked by their center of mass — the absolute global maximum for symmetric kernels. Therefore, it is possible to study their evolution through scales and to provide the scale-space trajectory of their center of mass as the solution of an Ordinary Differential Equation (ODE). This model drastically reduces the redundancy of the wavelet representation in correspondence to significant structures of the signal. It also preserves correlation between adjacent coefficients at each scale: their behaviour is predicted from finer to coarser scale without particular requirements on the structure of the original signal.

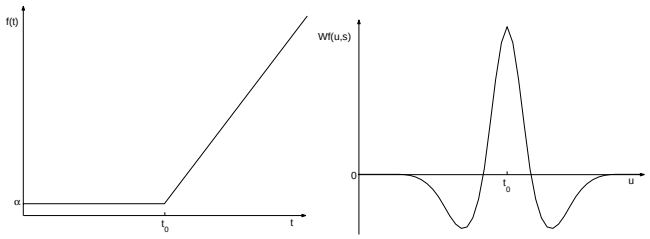

Figure 1: *Left)* Infinite ramp signal. *Right)* Its corresponding atom in the wavelet domain.

The paper is organized as follows. In Section 2 the atomic representation of the wavelet transform is briefly presented. Section 3 introduces the evolution law of wavelet atoms and yields their trajectories along scales. Section 4 shows how the atoms based model can be successfully used for de-noising. Finally, some experimental results will be given in Section 5 along with comparisons with the most recent and powerful denoising approaches.

2. WISDOW

In [1] a novel de-noising strategy based on an atomic representation of the wavelet transform of the signal has been proposed. A *basic atom* at scale s and centered at the location t_0 has been defined by the following function

$$F(t_0,u,s) = s\sqrt{s}\left(\int_{\frac{t_0-u}{s}}^{b} t\psi(t)dt - \frac{t_0-u}{s}\int_{\frac{t_0-u}{s}}^{b}\psi(t)dt\right) \quad (1)$$

where $u \in \mathbf{R}$ is the time variable (see Fig. 1) and ψ is the wavelet. It corresponds to the wavelet transform of an infinite ramp signal having the singularity located at t_0 with slope $\alpha_0 = 1$ and behaves as a wave that grows while dilating along scales.

The wavelet transform $Wf(u,s)$ of a generic signal f has been written as superimposition of basic atoms by exploiting

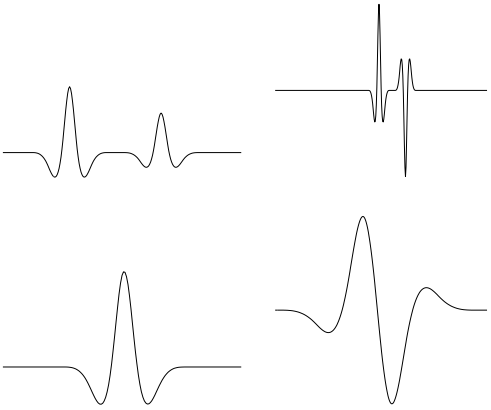

Figure 2: The two atoms on the top approach the corresponding bottomost one as the scale increases. *Left)* the atoms have the same sign; *Right*) the atoms have opposite sign.

the *overlapping effects principle* at each scale s. Therefore,

$$\forall\, s > 0, \quad Wf(u,s) \sim \sum_{k=1}^{N_s} \alpha_k F(t_k,u,s), \tag{2}$$

where α_k and t_k are respectively the slope and the location of each atom of the representation, while N_s is the number of the atoms used for the approximation at scale s. In fact, two atoms corresponding to two singularity points independently grow and dilate along scales till they begin to interfere and completely merge. From that point on, the shape of interfering atoms does not change while it still grows and dilates. In particular, if the atoms have the same sign, they converge to a single atom whose slope is the sum of the slopes of the two interfering atoms (see Fig.2.left). On the other hand, if the atoms have an opposite sign, they approach a wave which is characterized by two symmetric mainlobes having opposite sign (see Fig.2.right). It can be formally proved that a jump discontinuity is the superimposition of two interfering basic atoms having opposite sign and located at t_1 and t_2, when t_2 approaches t_1 [1]. On the other hand, the wavelet transform of a polynomial ramp is approximated by a basic atom and a residual:

$$Wf(u,s) = \alpha_0 F(t_0,u,s) + Q(u,s). \tag{3}$$

The residual $Q(u,s)$ is negligible if the terms of order greater than one in a linear approximation of $f(t)$ in the range $[t_0, t_0 + \Delta t]$ are negligible. It is obvious that the shift Δt is also strictly related to the wavelet support.

The atomic representation, as in eq. (2), enables to drastically reduce the information which is necessary for reconstructing the original signal. In fact, it consists of atoms locations t_k and their slopes α_k at each scale s. These two parameters contain information about the value of the wavelet transform of the signal at scale s within the support of the atom centered at t_k. Hence, it intrinsically preserves the correlation between adjacent coefficients of the wavelet decomposition without having a priori information about the analysed signal. This property can be successfully exploited in de-noising and allows to overcome some of the limits of standard approaches, like thresholding and attenuation. In fact, atoms slopes are estimated at each scale using a greedy algorithm on the amplitude of modulus maxima, i.e.

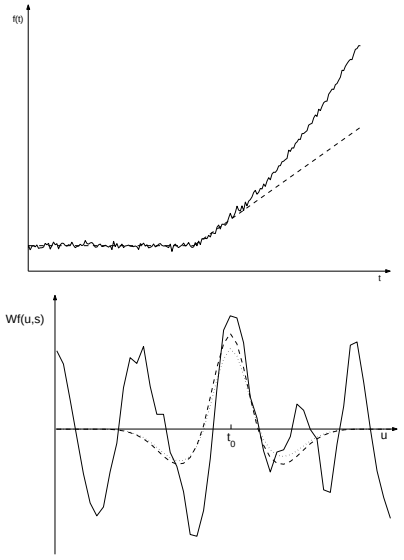

Figure 3: *Top)* Noisy polynomial signal (solid line), infinite ramp used for its approximation (dashed line). *Bottom)* Wavelet transform of: noisy polynomial signal (solid line), original signal (dotted line) and de-noised signal using the proposed approach (dashed line).

$$\alpha_k = \frac{< R_k(u,s), F(t_k,u,s) >}{\|F(t_k,u,s)\|^2}$$

where $R_k(u,s) = Wg(u,s) - \sum_{h=1}^{k-1} R_h(u,s)$. It corresponds to a least squares estimation in a suitable domain by imposing a function model and then providing a regularization of the noisy data (see Fig.3 and [1] for details). In [1] it has been rigorously proved that *WISDOW* formalism outperforms hard thresholding [6]. In fact, also coefficients under threshold are recovered, avoiding artifacts due to the rough cut off of information in the selection based approaches.

Even if WISDOW algorithm provides satisfying results in de-noising, it reveals some limits. The interference between atoms significantly affects the estimation of each single slope and the corresponding location. In fact, it is difficult to exactly separate each single contribution when two atoms are very close. Moreover, annoying spikes can affect the recovered signal since high amplitude noisy coefficients can be confused with real atoms. This drawback can be overcome by further characterizing the interaction between atoms through scales, as shown in the next section.

3. ATOMS EVOLUTION LAW

In this section we derive the trajectory through scales of the significant points of the wavelet transform of a signal. These points are the centers of mass of the atoms provided by the representation in eq. 2. In other words, from the atoms at scale s we want predict the locations of the atoms at scales greater than s. For the lack of space, we will omit formal calculations that can be found in [3].

For the infinite ramp function

$$f(t) = \begin{cases} \beta & t < t_1 \\ \alpha_1 (t - t_1) + \beta & t \geq t_1, \end{cases}$$

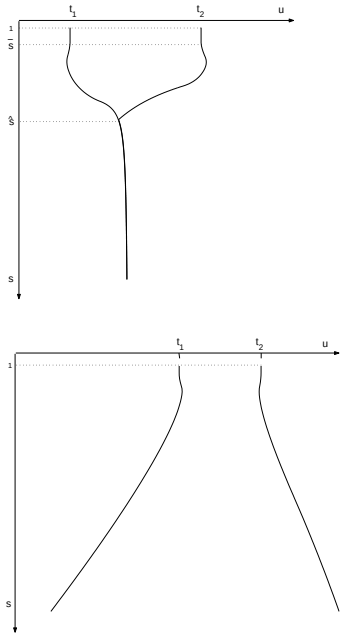

Figure 4: Atoms maxima trajectories along scales having the same sign (*top*) and opposite sign (*bottom*). They are the solutions of (6) with initial conditions respectively $u(s=1)=t_1$ and $u(s=1)=t_2$. $\overline{s}$ is the scale level at which interference begins while $\hat{s}$ is the one at which interference is complete.

it is possible to write the following evolution law

$$\frac{\partial}{\partial s}w = \frac{t_1-u}{s}\frac{\partial}{\partial u}w + \frac{3}{2s}w. \tag{4}$$

where $w(u,s) = Wf(u,s)$. It is a transport equation with a source term: the atom dilates along scales while its energy grows. The term $\frac{3}{2s}w$ measures the growing of the energy while the *characteristic curves* in the scale-space domain describe the dilation. In fact, the latter are continuous straight lines and give the trajectories of each point composing the atom. In particular, the center of mass of the atom does not move from its initial position $u=t_1$. On the contrary, the trajectory of the other wavelet coefficients belonging to the atom are straight lines whose slope is proportional to the distance from the center of mass at scale $s=1$.

As already outlined in the Introduction, it is easier to manage atoms via their absolute maximum than their center of mass for real world signals. In this case, atom extrema trajectories can be computed as shown in the Appendix of [2] and satisfy the following equation:

$$\dot{u} = -\frac{t_1-u}{s}. \tag{5}$$

For the global maximum it holds $u(s=1)=t_1$ and then the solution of (5) is $u(s)=t_1, \quad \forall\, s\in\mathbf{R}^+$. Therefore, the global maximum of an isolated atom does not move from its initial position.

Previous results can be generalized to interfering atoms [2]. To this aim, we consider a function having two singularity points yielding two atoms located at t_1 and t_2. In this case equation for atoms modulus maxima trajectories becomes:

$$\dot{u} = -\frac{\overline{t}-u}{s} - \frac{d}{2s}\left(\frac{\alpha_2\psi(z_2)+\alpha_1\psi(z_1)}{\alpha_2\psi(z_2)-\alpha_1\psi(z_1)}\right), \tag{6}$$

where $\overline{t}=\frac{t_1+t_2}{2}$, $\{z_k=\frac{t_k-u}{s}, \quad k=1,2\}$ and $d=t_2-t_1$.

Eq. (6) differs from (5) for its last non linear term. This latter describes how the two atoms move and interfere along scales. While there is no interference, the two cones of influence do not overlap and the two atoms are independent of each other. They grow and dilate but their global maxima do not move from their initial position (see the topmost picture in Fig. 4 for $s<\overline{s}$)). Once the two cones of influence begin to overlap, each atom is influenced by the other one. In particular, the atom with the greater amplitude reveals a stronger attractive force, causing a greater deviation in the trajectory of the smaller atom, as depicted in Fig. 4.top (for $\overline{s}<s<\hat{s}$) [1]. Whenever the interference is complete, i.e. for $s>\hat{s}$, the two atoms become an only one, as depicted in Fig. 2. It is worth outlining that (6) it is non reversible, i.e. at a fixed scale s it is not possible to reconstruct the history of a single atom without ambiguity. More precisely, at a fixed scale it is not possible to say whether an atom corresponds to an isolated singularity or it is the composition of two or more atoms.

The maxima trajectories for N interfering atoms is straightforward:

$$\dot{u} = -\frac{\overline{t}-u}{s} - \frac{1}{s}\frac{\sum_{k=1}^{N} d_k\alpha_k\,\psi\left(\frac{t_k-u}{s}\right)}{\sum_{k=1}^{N}\alpha_k\psi\left(\frac{t_k-u}{s}\right)}, \tag{7}$$

where $\overline{t}=\frac{\sum_{k=1}^{N}t_k}{N}$, $d_k=t_k-\overline{t}$.

Even in this case, the last non linear term measures the shift of atoms modulus maxima according to their distance and their amplitude.

3.1 About Lipschitz order of singularities

So far, we used a fixed atom shape for modelling various kinds of singularity. It is symmetric and corresponds to a singularity of a piecewise linear function $f(t)$. Its trajectory is described by (5) and the decay of its amplitude along scales is $O(s^{3/2})$.

Nonetheless, the amplitude of a wavelet modulus maximum is tied to the Lipschitz order of the corresponding singularity [9]. Therefore, its amplitude decays as $O(s^{\gamma+1/2})$. To overcome this problem we can generalize the atom equation as follows:

$$G(u,s) = \alpha_1 s^{\gamma-1}F(t_1,u,s), \tag{8}$$

where $F(t_1,u,s)$ is defined in (1). This way, the atom amplitude is modulated in agreement with the decay of the analyzed singularity while its shape is approximated with eq. (1).

Even in its general form, the atom again obeys to a precise evolution law from which it is possible to derive maxima trajectories. In fact, it holds

$$\frac{\partial}{\partial s}G = \frac{t_1-u}{s}\frac{\partial}{\partial u}G + \frac{\gamma+1/2}{s}G, \tag{9}$$

[1] It is worth highlighting that the two atoms show repulsion before attraction. It is due to the shape of each single atom which is not positive in its whole domain (see Fig. 1). It entails a shift in the modulus maxima locations whenever sidelobes overlap mainlobes.

and then its maximum trajectory is still $\dot{u} = -\frac{t_1-u}{s}$, since symmetric.

Finally, for N interfering atoms located at t_k with growing exponents γ_k, the atoms trajectories are then the solution of the following ode:

$$\dot{u} = -\frac{\bar{t}-u}{s} - \frac{1}{s}\frac{\sum_{k=1}^{N} d_k \alpha_k s^{\gamma_k} \psi\left(\frac{t_k-u}{s}\right)}{\sum_{k=1}^{N} \alpha_k s^{\gamma_k} \psi\left(\frac{t_k-u}{s}\right)} + \\ -\frac{1}{s}\frac{\sum_{k=1}^{N} \gamma_k \alpha_k s^{\gamma_k} \int_{\frac{t_k-u}{s}}^{+\infty} \psi(y)dy}{\sum_{k=1}^{N} \alpha_k s^{\gamma_k} \psi\left(\frac{t_k-u}{s}\right)}. \qquad (10)$$

The solution of the ODE (10) is determined by the initial conditions $\{t_k, \alpha_k, \gamma_k\}_{1\le k\le N}$, which respectively are the locations, the slopes and the decay exponents of atoms at $s = 1$. Hence, the knowledge of $\{t_k, \alpha_k, \gamma_k\}_{1\le k\le N}$ allows to predict wavelet coefficients at all successive scale levels. In Appendix we briefly describe how to determine these once initial conditions.

4. DENOISING

The evolution laws model both interscale and intrascale dependencies since they take into account the interaction between interfering atoms along scales: at each time we know the location and the state of each atom. This property allows us to overcome some of the limits of WISDOW denoising algorithm. In fact, the evolution laws enable to separate atoms contribution at a fixed scale, yielding a better estimation of their slopes. Therefore, for each domain of estimation we can predict how many contributions have to be considered along with their locations. The building of maxima chains enables to use also scale levels where atoms interference is complete without loosing important information. In that way we can exploit the fact that at coarsest scales the noise flattens since its negative Lipschitz order γ [9] while the estimation domains become wider, since the dilation property of the wavelet transform. It turns out that the least squares used for slopes estimation are more precise. It is worth to further outlining that atoms decay allows to discard those coefficients dominated by noise, since their negative γ.

4.1 The Algorithm

In this section the algorithm for restoring a noisy signal g is described in detail.

Let us consider an overcomplete wavelet decomposition [9] of g. The overcomplete representation is employed to avoid problems due to decimation. In fact, this operation unavoidably causes distortions of the shape in Fig. 1.

1. Perform the undecimated wavelet transform of g up to J^{th} scale level.
2. Perform the continuous wavelet transform on g at scales $s \in [1,2)$ using the step $\Delta s = .05$.
3. Estimate the parameters $\{t_k, \alpha_k, \gamma_k\}$ using WISDOW slope estimation algorithm three times for getting α_k, (as described at the end of the previous section) and t_k at scale $s = 1$ and (11) for estimatimg γ_k.
4. Eliminate atoms having $\gamma_k < 0$.
5. Compute atoms trajectories by solving (10) using a 4^{th} order Runge Kutta method and extract the solution at dyadic scales $s = 2^j, \quad j = 1,\ldots,J$.

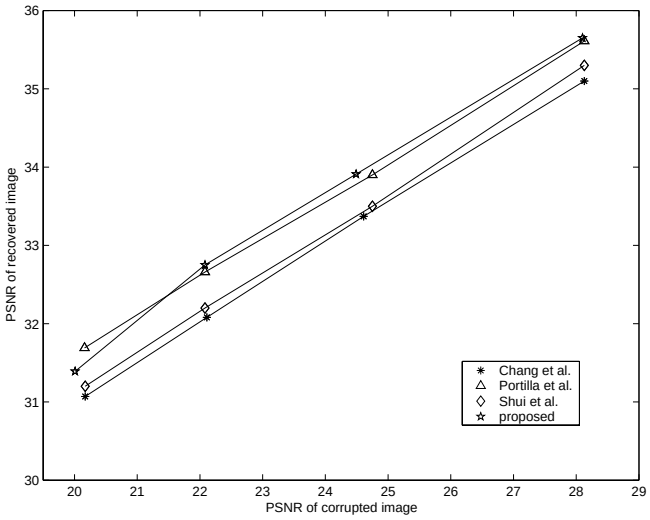

Figure 5: PSNR values versus noise variance for [10], [4], [11], and the proposed approach.

6. At scale $s = 2^j, \quad j = 1,\ldots,J$, sort selected maxima in decreasing order with respect to their absolute value and estimate the corresponding slope α_k. The data to use in the least squares estimation are weighted proportionally to the ratio between the analysed maximum and the ones in its cones of influence, which have been predicted by the law.
7. Invert the undecimated wavelet transform using the recovered detail bands.

5. EXPERIMENTAL RESULTS

The algorithm has been tested on many signals and images. A biorthogonal wavelet 3/9 associated to an over-complete multi-resolution decomposition has been adopted in all tests. Four scale levels have been used while we set the integration step $h = 0.05$ for solving the ode in the step 5 of the denoising algorithm.

Images are split into independent 1D signals. It is just a first attempt to use the evolution laws for images since there is not a formal extension to the 2D case. Nonetheless, experimental results show that the proposed denoising algorithm outperforms the most effective wavelet based denoising approaches. In particular, Fig. 5 compares the proposed model with the gaussian mixture estimator presented in [10], the adaptive bayesian thresholding using context modelling contained in [4] and the local Wiener filtering using elliptic directional windows for different subband in [11]. In fact, evolution law establishes a precise link between corresponding coefficients at different scales and allows us to well manage the interference between singularities even at coarser scales. This entails an almost faithful reconstruction of the original signal, avoiding constraints on the minimum distance between them, as in [7]. For a visual evaluation of the results, in Fig. 6 the denoised $512 \times 512 \times 8$ Lena image is depicted (PSNR = 32.75) — the noise level is $\sigma = 20$ (PSNR = 22.06). It is worth outlining that edges are well recovered thanks to the ability of the proposed scheme in modelling discontinuity points in the scale space domain yielding a drastic reduction of both ringing effect around edges and spikes — these latter

Figure 6: Denoised Lena image using the proposed approach (noisy image: PSNR = 22.06 db; recovered: PSNR = 32.75 db).

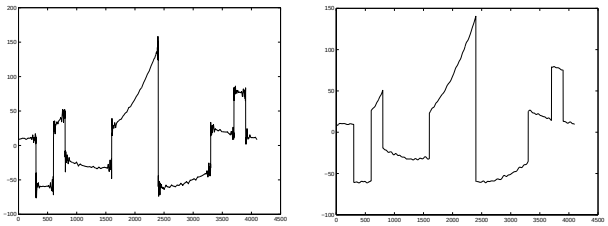

Figure 7: Recovered piecewise polynomial signal (noise SNR = 20.37 db) using: (left) Hard-thresholding (SNR = 26.88 db); (right) modified *WISDOW* (SNR = 32.38 db).

are due to isolated noisy coefficients. As an example, Fig. 7 shows the denoised piecewise polynomial test signal using the proposed approach and compares it with the recovered one using hard-thresholding.

The main computational effort of the algorithm is required by the solution of the ode in (10). Experimental results show that for getting acceptable prediction of atoms locations at dyadic scales the integration step h has to be lower than 0.1.

A. INITIAL CONDITIONS

α_k and t_k can be estimated at $s = 1$ using the atoms estimation algorithm used in WISDOW [1] while the decay exponents require a further analysis. They can be estimated by solving (9) in a suitable interval $[1, 1+\Delta s]$. Δs has to be quite small for guaranteeing that the interference between atoms does not still affect the locations of their maxima. Under this assumption, each atom can be considered isolated and then the equation (9) can be solved for each of them, yielding

$$G_s^{(k)} = \frac{\gamma_k + 1/2}{s} G^{(k)},$$

with the following initial condition $G^{(k)}(u(1),1) = \alpha_k F(t_0, t_0, 1)$. Hence, $G^{(k)} = C_0 s^{\gamma_k+1/2}$, that is $\alpha_k(s) s\sqrt{s} = \alpha_k(1) s^{\gamma_k+1/2}$, and then

$$\gamma_k = \log_s(\alpha_k(s)/\alpha_k(1)) + 1. \tag{11}$$

It is worth outlining that the estimation of γ_k depends on the estimation of the corresponding slope α_k. The better α_k estimation the better more faithful γ_k value. To this aim we can iterate the algorithm used for slopes estimation. More precisely, if $E(u,s)$ is the error for the atomic approximation at a fixed scale s, we can iterate the decomposition algorithm on it and combine the results with the ones found at the previous step.

REFERENCES

[1] V. Bruni and D. Vitulano, "Wavelet based Signal Denoising via Simple Singularities Approximation," to appear in *Signal Processing Journal* Elsevier Science.

[2] V. Bruni, B. Piccoli, and D. Vitulano, "Wavelet time-scale Dependencies for Signal and Image Compression," *Proc. ISPA 2005*, Zagreb, Croatia, September 15-18. 2005, pp. 105-106.

[3] V. Bruni, B. Piccoli, and D. Vitulano, "Scale space atoms for signals and image de-noising," *IAC-CNR Report*, 2006.

[4] S.G. Chang, Bin Yu and M. Vetterli, "Spatially Adaptive Thresholding with Context Modeling for Image Denoising", *IEEE Transactions on Image Processing*, Vol. 9, No. 9, pp. 1522-1530, September 2000.

[5] M. N. Do and M. Vetterli, " Contourlets: A New Directional Multiresolution Image Representation", *Proc. 36th Asilomar Conf. on Signals Systems and Computers*, Vol.1, November 2002, pp. 497-501.

[6] D. L. Donoho and I. M. Johnstone, "Ideal Spatial Adaptation via Wavelet Shrinkage", *Biometrika*, Vol. 81, pp. 425-455, 1994.

[7] P.L. Dragotti and M. Vetterli, " Wavelet Footprints: Theory, Algorithms and Applications", *IEEE Transactions on Signal Processing*, Vol. 51, No. 5, pp. 1306-1323, May 2003.

[8] E. Le Pennec and S. Mallat, " Non linear image approximation with bandelets", *Tech. Rep. CMAP/ Ecole Polytechnique*, 2003.

[9] S. Mallat, *A Wavelet Tour of Signal Processing*, Academic Press, 1998.

[10] J. Portilla, V. Strela, M. Wainwright and E. Simoncelli, " Image Denoising using Scale Mixtures of Gaussians in the Wavelet Domain", *IEEE Transactions on Image Processing*, Vol. 12, No. 11, pp. 1338-1351, November 2003.

[11] P. L. Shui, " Image Denosing Algorithm via Doubly Local Wiener Filtering with Directional Windows in Wavelet Domain" *IEEE Signal Processing Letters*, Vol. 12, No. 10, pp. 681-684, October 2005.

[12] J. L. Starck, E. J. Candes and D. L. Donoho, " The Curvelet Transform for Image Denoising " *IEEE Transactions on Image Processing*, Vol. 11, No. 6, pp. 670-684, June 2002.