J Math Imaging Vis (2012) 44:52–64
DOI 10.1007/s10851-011-0310-2

Time-Scale Similarities for Robust Image De-noising

Vittoria Bruni · Domenico Vitulano

Published online: 16 August 2011

Abstract This paper presents a novel image denoising algorithm, namely *Atomic Non Local Means* (ANL-means), that looks for similarities in the time-scale domain. To this aim, wavelet details are approximated by linear combinations of predefined atoms, whose centers of mass trace trajectories in the time-scale plane (from fine to coarse scales). These trajectories depend on the mutual distance between not isolated singularities, their different decay along scales and their amplitude ratio. These three parameters have proved to be useful in catching image self-similarities and in the implementation of a robust NL-means based denoising algorithm. *ANL-means* is able to reach and often outperform the most powerful and recent NL-means based de-noising schemes in terms of both mean square error and visual quality.

Keywords Image denoising · NL-means · Wavelets · Atomic approximation · Time-scale analysis

1 Introduction

De-noising is a fascinating and widely investigated topic of image processing, as proved by the significant related literature—see [1–18] and references therein. It is often used as an essential preprocessing step in many applications, since it aims at recovering a signal f from its corrupted version g, that is supposed to be

$$g(t) = f(t) + \upsilon(t), \quad t \in \mathbf{R}, \tag{1}$$

with υ additive zero mean Gaussian noise with variance σ^2.

The solution of this problem often requires the definition of models for the original image that are able to emphasize its spatial and frequency correlations, that are opposite to the uncorrelated nature of the noise. The big effort devoted to define more or less sophisticated regularization methods (anisotropic smoothing, variational models [13, 15, 19], etc.) and the construction of effective expansion bases (yielding a sparse image representation) [9, 20–22] provided very interesting and impressive denoising results. Nonetheless, those methods only depend on local spatial variations of the signal, at the same or at different resolutions [12, 16, 17].

The new trend is to exploit not only the local spatial and frequency redundancy, but also the non local image self similarities. Similar objects may not be necessarily spatially close, but they can occupy different locations in the image. Thus, a more promising solution for denoising is the use of this correlated redundant information in contrast with the random nature of the noise. This is what Buades, Coll and Morel [4] proposed in their work: each pixel is cleaned by means of a weighted average that involves all image pixels; the weights account for the similarity between luminance values—the more similar these values, the higher the weight. Pixel similarity is measured in terms of l^2 norm of their neighborhood luminance values: the closer their neighborhood, the more similar the pixels. This method (called Non Local Means, or simply NL-means), improves classical denoising approaches by using the Lebesgue measure in conjunction with the Riemannian one, as formally proved in [18]. However, its main drawbacks are a prohibitive computing time, the lack of denoising whenever similarities are

V. Bruni (✉)
Faculty of Engineering—Dept. of SBAI, University of Rome La Sapienza, Via A. Scarpa 16, 00161 Rome, Italy
e-mail: bruni@dmmm.uniroma1.it

D. Vitulano
Istituto per le Applicazioni del Calcolo "M. Picone"—C.N.R., Via dei Taurini 19, 00185 Rome, Italy
e-mail: vitulano@iac.rm.cnr.it

not found and the choice of the best domain for catching similarities. In order to overcome these limits, various attempts have been proposed in the literature. For example, some works make NL-means faster by reducing the search domain eliminating insignificant pixels or by optimizing the computation of the similarity measure [1, 13, 23, 24]. Most of them also use a different domain where to catch similarities. In fact, the simple l^2 norm of the neighborhood vectors may not be adequate while more proper neighborhood features assure a more reliable similarity measure. For instance, Coupé et al. implemented a block version of NL-means by mixing wavelet subbands [25], Wang et al. [26] proposed a fast computation of similarities through the fast Fourier transform, Mahmoudi et al. [27] used a pre-classification of pixels neighborhood, Azzabou et al. in [1] applied NL-means by computing an adapted dictionary while Kervrann et al. in [13] introduced local dictionaries as well as a Bayesian distance measure (*Bayesian NL-means*). On the other hand, in order to improve denoising performance, Foi et al. [6] used block matching to find similarities and a 3D transform to clean the found correlated blocks, Tasdizen [28] used the Principal Component Analysis to define a subspace where to catch pixel similarities, Chatterjee et al. [29] modeled the NL-means as a 0th order kernel regression method, Brox et al. [30] embedded NL-means in a variational formulation oriented to texture restoration, Van De Ville et al. [31] derived the SURE expression for NL-means algorithm, while Goossens et al. [32] proved that the use of measures different from the simple l^2 norm (Leclerc measure) provides better results even in case of correlated noise.

The neighborhood size in an NL-means based algorithm, the smoothing percentage parameter as well as the setting of the best subspace for a quick search of similarities are strictly linked to the multiscale information of the analyzed image. For that reason, we employ a time-scale representation of the image for gathering a proper set of parameters (features) to use in the evaluation of image similarities. In particular, we aim at characterizing the multiscale behavior of different points in the image to assess pixel similarity. It is equivalent to simultaneously using neighborhoods with different size. Moreover, the transformed domain also emphasizes the relation of each pixel with its neighborhood. Wavelet time-scale representation has been adopted in this paper due to its ability in characterizing the regularity of edge points through the decay of the amplitude of the corresponding wavelet coefficients along scales [14, 33]. In particular, the properties of the atomic approximation in [34] are very appropriate for the final task. The latter approximates wavelet details through interacting atoms (i.e. groups of adjacent coefficients) that define interesting geometric objects in the time scale plane. They are the trajectories of atoms modulus maxima. These trajectories account for the location and amplitude of significant coefficients in the time scale plane that correspond to significant structures of the image under study. Hence, similar objects in the image provide similar atoms trajectories in the time-scale plane, namely points belonging to the same edge and points belonging to the same type of edge but located in a different area of the image. Such an approach, called *ANL-means* (Atomic Non Local means), automatically sets the block size in which to find the similarity parameters: it corresponds to the atoms support at each level of resolution. Moreover, it provides denoising even if similarities are missing, thanks to the atomic approximation properties. Finally, it gives the second dimension to the monodimensional atomic representation (close points of the same edge are similar with respect to the employed measure). Extensive experimental results and comparative studies show the effectiveness of the proposed denoiser from both objective (PSNR) and subjective (visual quality) point of view.

The outline of the paper is the following. Section 2 briefly presents the properties of the atomic approximation and provides a similarity measure for wavelet coefficients belonging to the same or similar edge, as well as its use for image denoising. Finally, comparative studies, discussions and experimental results are offered in Sect. 3.

2 Time-Scale Similarities

The aim of this section is to characterize the neighborhood of each pixel through its multiscale behavior in a wavelet decomposition. The goal is to define specific features of the neighborhood to use for the evaluation of proper similarity measures.

Let us consider two step functions characterized by different gray values but the same step amplitude. The comparison of the gray levels of the neighborhood of the two discontinuity points characterizes them as different singularities. On the contrary, their neighborhoods result the same with respect to the first derivative of the two step functions (or equivalently their decomposition in a wavelet basis). This simple example shows that a successful comparison between different objects requires a proper features space. Let us now consider the two signals in Fig. 1. They are two distinct piecewise linear signals having the same wavelet detail coefficients. In particular, the locations of significant modulus maxima of their wavelet transform belong to the same curves in the time scale plane. These signals have different low-pass component while the same details (high-pass component). Therefore, it is more convenient to find similarities among wavelet coefficients and then to apply them NL-means filter.

In the following section the wavelet transform of a signal is written as superposition of traveling simple waves in the

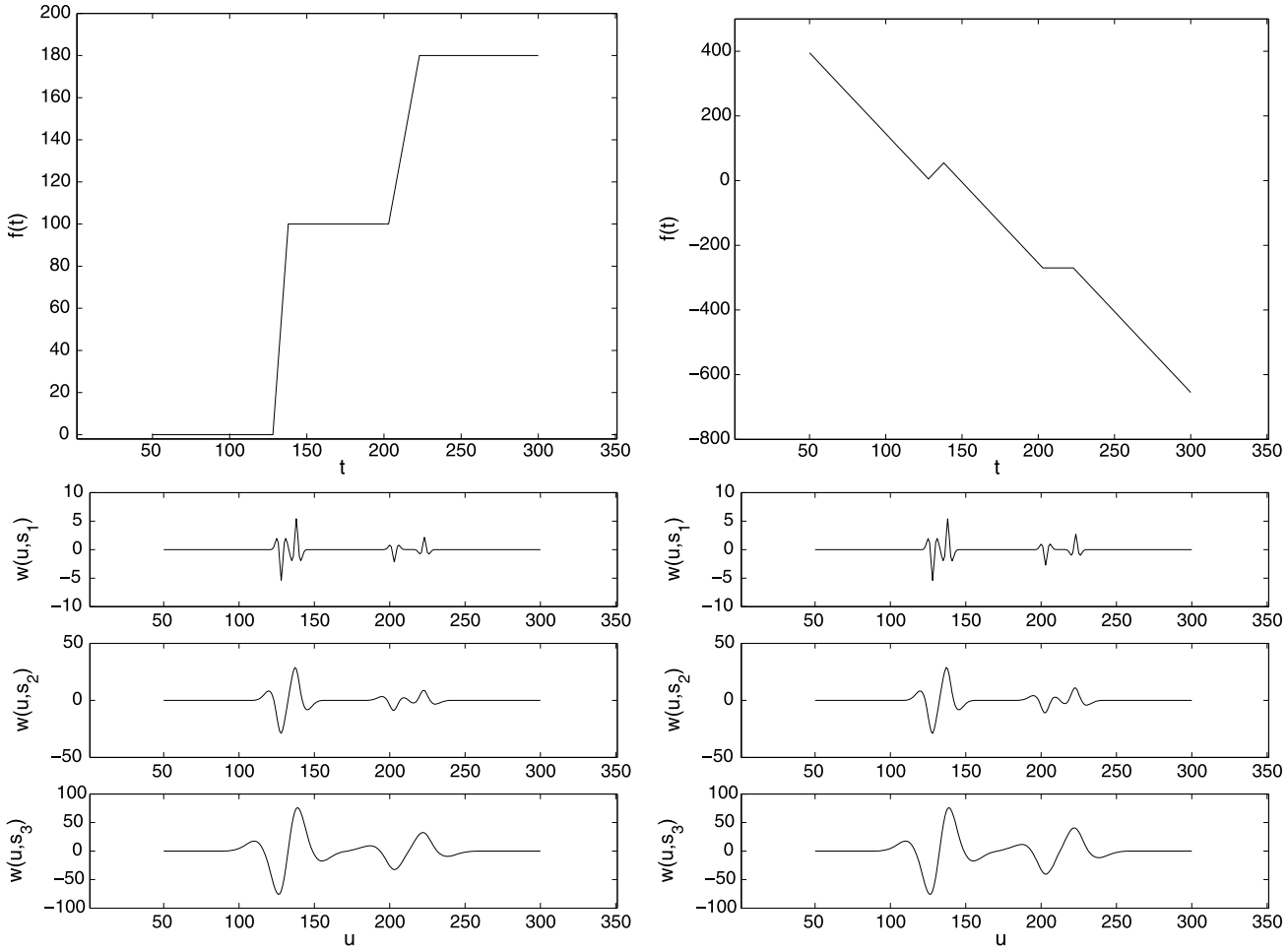


Fig. 1 (*Top*) Two different piecewise linear signals. (*Bottom*) Their continuous wavelet transform computed at selected scales

time-scale (u, s) plane, that will be called *basic atoms*. Each atom corresponds to a singularity in the time domain. The trajectories in the time-scale plane of atoms global maxima describe the interaction between neighboring singularities. Thus, similar edges are those providing the same modulus maxima curves in the time scale plane.

2.1 Atoms Trajectories in the Time Scale Plane

The wavelet transform $w(u, s)$ of a function f can be approximated through a linear combination of predefined atoms [34] as follows

$$w(u,s) \sim \sum_{k=1}^{N} \alpha_k s^{\gamma_k - 1} F(t_k, u, s), \quad u \in \mathbf{R}, s \in \mathbf{R}^+, \tag{2}$$

with

$$w(u,s) = \frac{1}{\sqrt{s}} \int_{\frac{t_k-u}{s}}^{+\infty} f(t) \psi \left(\frac{t_k - u}{s} \right) dt,$$

$$F(t_k, u, s) = s\sqrt{s} \left\{ \int_{\frac{t_k-u}{s}}^{+\infty} t\psi(t)dt - \left(\frac{t_k - u}{s} \right) \int_{-\infty}^{+\infty} \psi(t)dt \right\}, \tag{3}$$

$$\alpha_k = \frac{\langle F(t_k, u, s), w(u, s) \rangle}{s^{\gamma_k - 1} \| F(t_k, u, s) \|^2}, \tag{4}$$

and ψ is a symmetric and compactly supported wavelet with support $[-C, C]$. $F(t_k, u, s)$ is called *basic atom* and it is depicted in Fig. 2. It corresponds to the wavelet transform of a linear ramp signal with unitary slope, whose first derivative is discontinuous at t_k. Each atom has an extremum point at t_k that corresponds to its center of mass. For this point it is possible to build the trajectory in the time-scale plane (u, s). The following proposition proves that it is possible to derive the trajectories $\overline{u} = u(s)$ in the (u, s) plane of the extrema of $w(u, s)$.

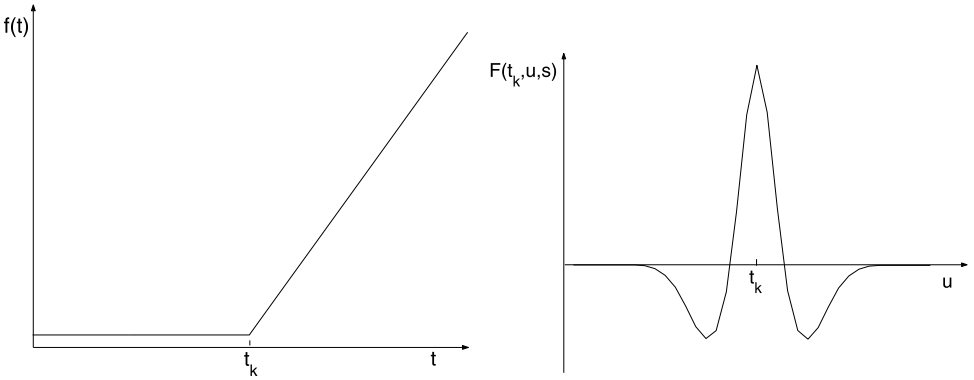


Fig. 2 Wavelet transform of a linear ramp signal having a singularity point at t_k, computed at a fixed scale s using a 3/9 spline biorthogonal wavelet. It is the *basic atom* in (3)

Proposition 1 *The trajectories $\overline{u} = u(s)$ of the extrema of $w(u,s)$ in (2) are the solutions of the following equations*

$$\begin{cases} \dot{u} = -\frac{t_k-u}{s} - \left(\frac{1}{s}\sum_{h=1}^{N} \alpha_h s^{\gamma_h} d_{kh}\psi\left(\frac{t_h-u}{s}\right)\right) \\ \qquad / \sum_{h=1}^{N} \alpha_h s^{\gamma_h}\psi\left(\frac{t_h-u}{s}\right) \\ \qquad - \left(\sum_{h=1}^{N} \alpha_h \gamma_h s^{\gamma_h} \int_{\left(\frac{t_h-u}{s}\right)}^{+\infty} \psi(y)dy\right) \\ \qquad / \sum_{h=1}^{N} \alpha_h s^{\gamma_h}\psi\left(\frac{t_h-u}{s}\right), \\ u(s_1) = t_k, \quad k = 1,\ldots,N \end{cases} \tag{5}$$

where $d_{kh} = t_h - t_k$ and s_1 is the initial scale point.

Proof The key point of the proof is the evolution law for $w(u,s)$ in the (u,s) plane and its restriction over w's extrema. $w(u,s)$ in (2) can be rewritten as $w(u,s) = \sum_{k=1}^{N} w^{(k)}(u,s)$, where $w^{(k)}(u,s) = \alpha_k s^{\gamma-1} F(t_k,u,s)$. A direct comparison between the first order partial derivatives of w with respect to the variables u and s, respectively $w_u(u,s)$ and $w_s(u,s)$, provides

$$w_s = \frac{t_k-u}{s} w_u + \frac{1+2\sum_{h=1}^{N}\gamma_h}{2s} w + \frac{1}{s}\sum_{h=1}^{N} d_{kh} w_u^{(h)} + \frac{1}{s}\sum_{h=1}^{N}\gamma_h w^{(h)}, \tag{6}$$

where $d_{kh} = t_h - t_k$.

Let now $\overline{u} = u(s)$ be a maxima curve in the (u,s)-plane, then $w_u(\overline{u},s) = 0$ and

$$\left((w(s,u(s))_u\right)_s = 0,$$

that is

$$w_{su} + w_{uu}\dot{u} = 0, \tag{7}$$

where $\dot{u}$ is the derivative of $u(s)$ with respect to s. On the other hand, the partial derivative of both members of (6) in correspondence to $\overline{u} = u(s)$ gives

$$w_{su} = \frac{t_k-u}{s} w_{uu} + \frac{1}{s}\sum_{h=1}^{N} d_{kh} w_{uu}^{(h)} + \frac{1}{s}\sum_{h=1}^{N}\gamma_k w_u^{(h)}, \tag{8}$$

where

$$w_u^{(h)} = \alpha_h s^{\gamma_h - \frac{1}{2}} \int_{\frac{t_h-u}{s}}^{+\infty} \psi(y)dy,$$

$$w_{uu}^{(h)} = \alpha_h s^{\gamma_h - \frac{3}{2}} \psi\left(\frac{t_h-u}{s}\right).$$

Equation (5) derives from a direct comparison between (8) and (7). □

Let $\Omega_h = \{u : |u - t_h| \le Cs\}$ be the cone of influence of the singularity at t_h, $\forall h$. Then, if the atom at t_k is isolated, i.e. $\Omega_k \cap \Omega_h = \emptyset\ \forall h \ne k$ at each scale s, the last two terms of the second member of (5) are zero since $\psi\left(\frac{t_k-u(s)}{s}\right) \ne 0 \Leftrightarrow u : |u - t_k| \le Cs$. Thus, $u(s) = t_k$, $\forall s$ i.e., the extremum point in t_k does not move from its original location. Whenever the cones of influence of two atoms intersect, the location of their corresponding global maxima changes according to the quantity

$$\frac{\sum_{h=1}^{2} \alpha_h s^{\gamma_h} d_{kh}\psi\left(\frac{t_h-u}{s}\right)}{\sum_{h=1}^{2} \alpha_h s^{\gamma_h}\psi\left(\frac{t_h-u}{s}\right)} + \frac{\sum_{h=1}^{2} \alpha_h \gamma_h s^{\gamma_h} \int_{\left(\frac{t_h-u}{s}\right)}^{+\infty} \psi(y)dy}{\sum_{h=1}^{2} \alpha_h s^{\gamma_h}\psi\left(\frac{t_h-u}{s}\right)}.$$

More precisely, if α_1 is the smallest atom, its shift is regulated by the quantity

$$\frac{\alpha_2 s^{\gamma_2} d_{12}\psi\left(\frac{t_2-u}{s}\right)}{\sum_{h=1}^{2} \alpha_h s^{\gamma_h}\psi\left(\frac{t_h-u}{s}\right)} + \frac{\alpha_2\gamma_2 s^{\gamma_2} \int_{\left(\frac{t_2-u}{s}\right)}^{+\infty} \psi(y)dy}{\sum_{h=1}^{2} \alpha_h s^{\gamma_h}\psi\left(\frac{t_h-u}{s}\right)},$$

whose numerator is regulated by the largest one α_2, and vice versa. Then, the atom with the smallest slope and slower decay moves from its original location more quickly than the atom with the largest amplitude. In particular, if α_1 and α_2 have different sign, the two interfering atoms are subjected to a repulsive effect, otherwise there is attraction.

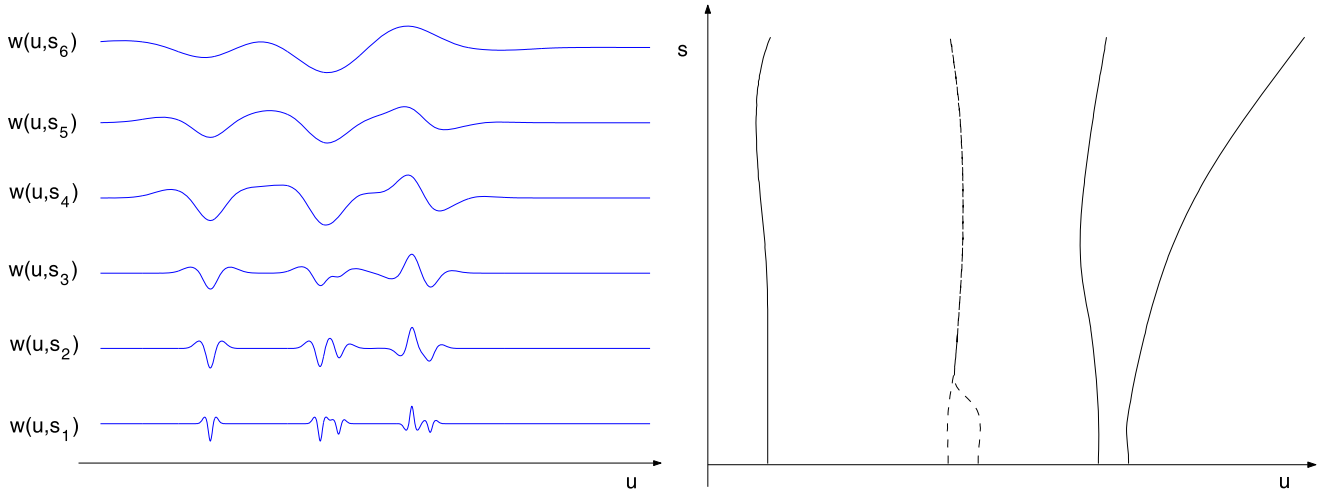


Fig. 3 Wavelet transform at selected scales of a signal composed of five atoms. The 2nd and 3rd atoms have the same sign, the last two have opposite sign (*left*). The first pair shows attraction while the last one repulsion (*right*). In both cases the smaller atom moves more quickly from its original location, according to (5)

Figure 3 depicts the trajectories of a signal composed of five atoms. The second and third ones have the same sign and the smaller one converges to the largest one till they create a single atom. At higher scales it is attracted by the isolated atom in the left side (same sign) while it is repulsed by the first atom in the right side (opposite sign). The latter is, in its turn, repulsed by the rightmost atom.

Hence, for a fixed k, atom's trajectory is influenced by atom's neighborhood that is composed of interfering atoms t_h such that $\Omega_k \cap \Omega_h \neq \emptyset$ for some set of subsequent scales. This is the neighborhood with respect to measure image similarities, while atoms slopes α_k are the quantities to be cleaned, according to (2).

2.2 Denoising Using Self-similarities

Equation (5) attempts to formalize *inter*- and *intra*-scale dependencies of wavelet coefficients by properly representing them. Those dependencies regulate the interaction between time-scale atoms. For a fixed k, by dividing both the numerator and the denominator of the last two terms of the second member of (5) by the quantity $\alpha_k s^{\gamma_k}$ we have:

$$\begin{cases} \dot{u} = -\frac{t_k-u}{s} - \left(\sum_{h=1}^{N} \frac{\alpha_h d_{kh} s^{\gamma_h-\gamma_k}}{s\alpha_k} \psi\left(\frac{t_h-u}{s}\right)\right) \\ \quad \Big/ \sum_{h=1}^{N} \frac{\alpha_h s^{\gamma_h-\gamma_k}}{\alpha_k} \psi\left(\frac{t_h-u}{s}\right) \\ \quad - \left(\sum_{h=1}^{N} \frac{\alpha_h \gamma_h s^{\gamma_h-\gamma_k}}{\alpha_k} \int_{\frac{t_h-u}{s}}^{+\infty} \psi(y)dy\right) \\ \quad \Big/ \sum_{h=1}^{N} \frac{\alpha_h s^{\gamma_h-\gamma_k}}{\alpha_k} \psi\left(\frac{t_h-u}{s}\right), \\ u(s_1) = t_k, \quad k = 1, \ldots, N. \end{cases} \tag{9}$$

Hence, the trajectory in the time-scale plane of the atom in t_k depends on its distance d_{kh} from neighboring atoms, the ratio between their amplitudes $\frac{\alpha_h}{\alpha_k}$ and the difference of their decay $\gamma_h - \gamma_k$. These three parameters are the atom's context to be used for assessing atoms similarities in the whole image domain. Figure 4 depicts a simple piecewise smooth signal. Its wavelet transform is composed of 7 atoms. The ones in t_1 and t_6 have the same context, since they correspond to the same type of singularity. On the contrary, the atom in t_7 has a different context that is closer to the one in t_4. More formally, let $\{\mathcal{A}_k\}_{1\leq k\leq N}$ be the set of the estimated atoms at scale $s = s_1$. Each $\mathcal{A}_k$ is characterized by the three parameters t_k, α_{k,s_1} and γ_k, i.e. location, amplitude and decay.

Definition 1 The neighborhood $\mathcal{N}_k$ of each atom $\mathcal{A}_k$ is composed of

$$\{\mathcal{A}_{k-n}, \ldots, \mathcal{A}_{k-1}, \mathcal{A}_{k+1}, \ldots, \mathcal{A}_{k+m}\},$$

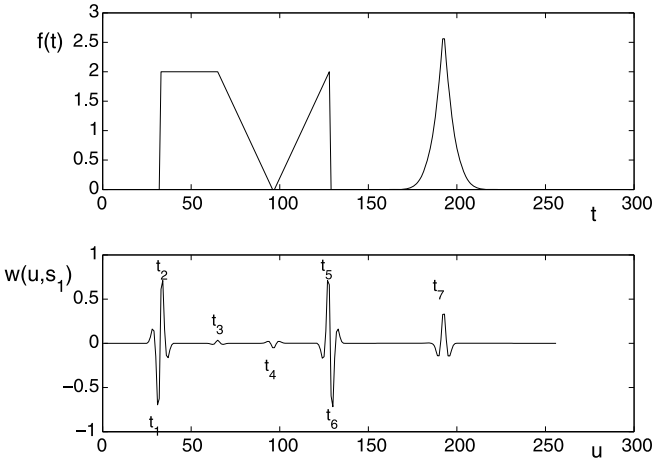
where n, m are such that $|t_{k-n} - t_k| \leq Cs_1$ and $|t_{k+m} - t_k| \leq Cs_1$.

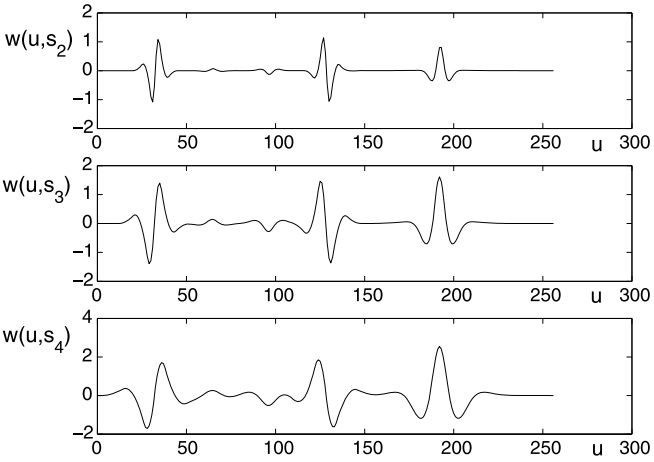
The elements belonging to $\mathcal{N}_k$ are the atoms that interfere with $\mathcal{A}_k$ from scale s_1 on. Hence, the distance between two atoms $\mathcal{A}_k$ and $\mathcal{A}_l$ is the distance between their neighborhood, i.e.

$$\|\mathcal{N}_k - \mathcal{N}_l\|_2^2.$$

Let V_k and V_l be the feature vectors of $\mathcal{N}_k$ and $\mathcal{N}_l$; then, neighborhood distance can be replaced by the distance between their feature vectors i.e.,

$$D_{kl} = \|V_k - V_l\|_2^2. \tag{10}$$





Atom	Location	Amplitude	Decay
$\mathcal{A}_1$	31	−2	0.009
$\mathcal{A}_2$	33	2	0.008
$\mathcal{A}_3$	64	0.08	1.42
$\mathcal{A}_4$	99	0.08	1.16
$\mathcal{A}_5$	127	2.02	0.007
$\mathcal{A}_6$	129	−2.02	0.0008
$\mathcal{A}_7$	192	0.57	1.34

$\mathcal{N}_1 = \{\mathcal{A}_2\} = \{(33, 2, 0.008)\}$
$\mathcal{N}_2 = \{\mathcal{A}_1\} = \{(31, -2, 0.009)\}$
$\mathcal{N}_3 = \mathcal{N}_4 = \mathcal{N}_7 = \{\emptyset\}$
$\mathcal{N}_5 = \{\mathcal{A}_6\} = \{(129, -2.02, 0.008)\}$
$\mathcal{N}_6 = \{\mathcal{A}_5\} = \{(127, 2.02, 0.007)\}$

Fig. 4 (*Topleft*) Piecewise regular signal and its wavelet transform at a fixed scale s_1. It is composed of seven atoms. (*Topright*) The wavelet transform of the same signal at scales larger than s_1. (*Bottom*) The table contains the parameters (location, amplitude and decay) and neighborhood of each atom

Specifically, V_k and V_l are the matrices whose elements define atom's context i.e.,

$$V_k = \begin{pmatrix} \frac{\alpha_{k-n,s_1}}{\alpha_{k,s_1}} & \gamma_k - \gamma_{k-n} & d_{k,k-n} \\ \vdots & \vdots & \vdots \\ \frac{\alpha_{k+m,s_1}}{\alpha_{k,s_1}} & \gamma_k - \gamma_{k+m} & d_{k,k+m} \end{pmatrix} \tag{11}$$

and V_l is defined in the same manner. V_k and V_l contain the initial conditions for (9), that describes the trajectories of the atoms $\mathcal{A}_k$ and $\mathcal{A}_l$ for some $k, l \in \{1, 2, \ldots, N\}$ in the time-scale plane.

The distance in (10) using feature matrices in (11) measures image similarities.[1] It combines the information contained at different resolutions through the decays γ_k, the set of interfering atoms (whose cardinality can change as the scale increases) and an adaptive neighborhood, that corresponds to atom support at the considered scales. Moreover, each slope $\alpha_{k,s}$ already embeds the relations between neighboring spatial pixels. It is a weighted average of the wavelet coefficients belonging to the cone of influence of the analyzed atom (see (4)).

In the presence of noise, the distances in (10) are suitable weights to use in the regularization of atoms slopes of the wavelet transform of the noisy data g. More precisely, the estimated slopes $\{\alpha_{k,s}\}_{1\le k\le N}$ at scale s can be corrected as follows

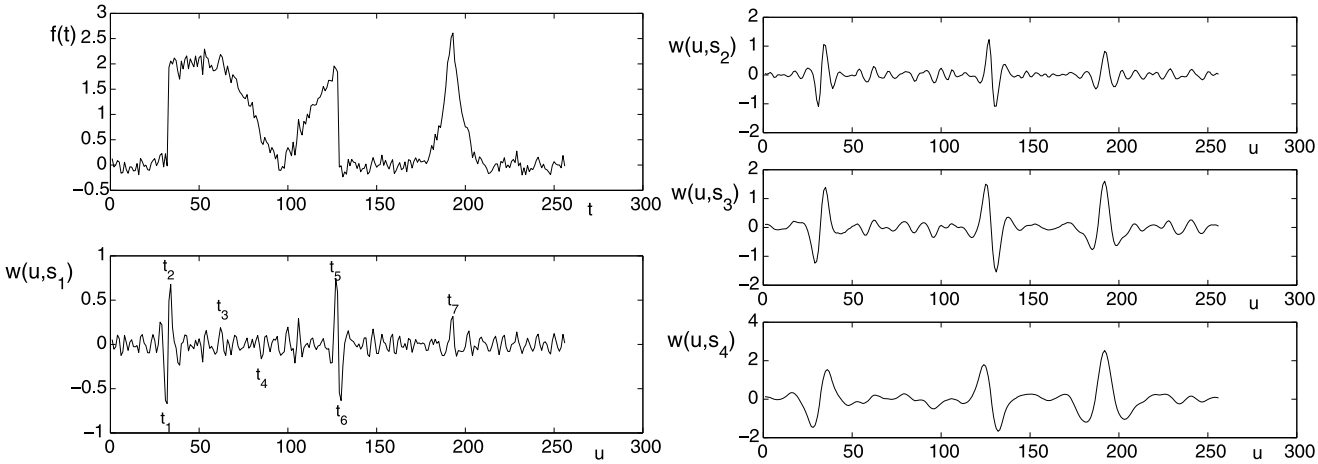
$$\overline{\alpha}_{k,s} = \frac{1}{\sum_{l=1}^{N} \omega_{kl}} \sum_{l=1}^{N} \omega_{kl}\alpha_{l,s}, \quad \forall k = 1, \ldots, N \tag{12}$$

with

$$\omega_{kl} = e^{-D_{kl}}. \tag{13}$$

The smaller the distance D_{kl} the larger the weight ω_{kl} (it approaches 1). In this way, the coefficients corresponding to the same class of singularities greatly contribute to the weighted average in (12). On the contrary, since wavelet coefficients corresponding to noise provide negative γ_k, they are discarded from the set of admissible initial conditions. Figure 5 depicts a simple signal corrupted with Gaussian noise. As it can be observed, singularity similarities are preserved even for increasing noise variance.

[1] If the neighborhoods of $\mathcal{A}_k$ and $\mathcal{A}_l$ have different cardinality, the smaller is zero-padded. Notice that the neighborhood of each atom is defined starting from the atom location and then symmetrically moving away from it toward the left and the right side. It is worth noticing that the different cardinality of the two neighborhoods could be due to a different local image content around the analyzed atoms or to a different influence of the noise (the same kind of singularity but with different amplitude).



	Without noise	With noise ($\sigma = .1$)	With noise ($\sigma = .3$)
$\omega_{1,5}$	.8687	.7912	.7201
$\omega_{1,6}$	.9077	.8533	.8013
$\omega_{1,4}$	.0001	.0075	.0083
$\omega_{1,7}$	.0001	.0011	.0010
$\omega_{3,4}$	.9182	.7701	.6558
$\omega_{4,7}$	.4434	.3922	.3201

Fig. 5 (*Topleft*) Piecewise regular signal in Fig. 4 corrupted by zero-mean Gaussian noise with standard deviation $\sigma = .1$ and its wavelet transform at a fixed scale s_1. (*Topright*) The wavelet transform of the same signal at scales greater than s_1. As the table of weights shows, the similarity between atoms is preserved in the presence of noise

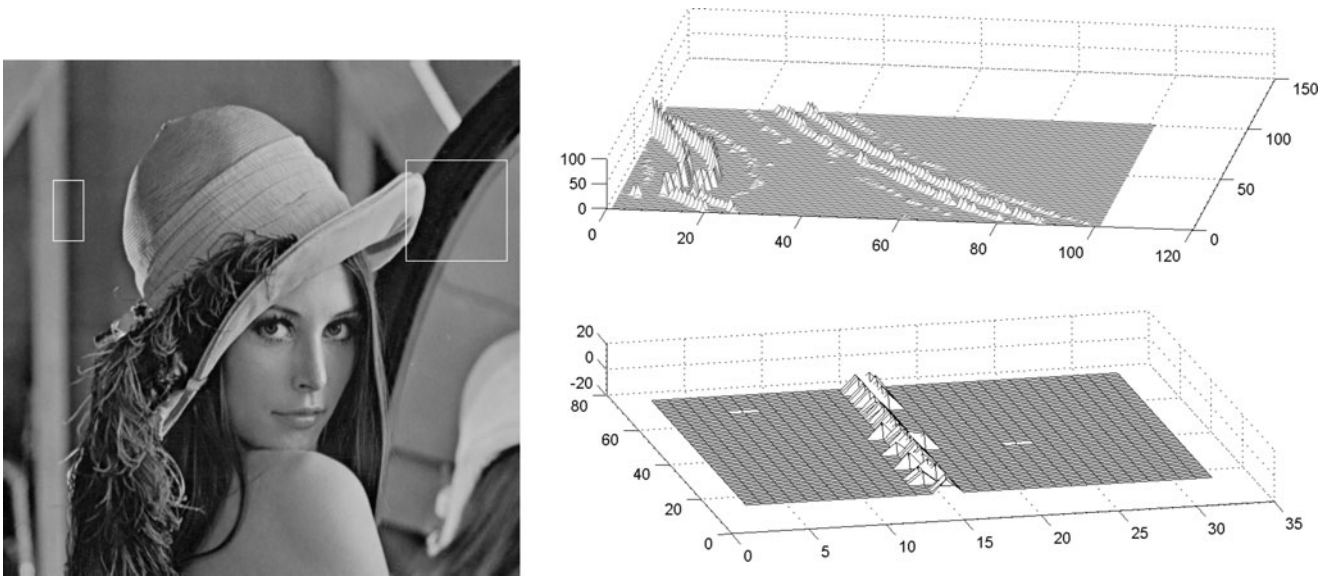


Fig. 6 Atoms distribution inside two regions (*white blocks*) of Lena image containing edge curves. The *topmost* mesh is relative to the largest block, while the *bottommost* mesh corresponds to the smallest block

It is worth noticing that if the initial scale point in (9) is updated, the distance in (10) can be evaluated just for a limited subset of scales, $I_s = [s_1, s_f]$. It turns out that the degree of similarity between two regions of an image can be measured at different resolutions. In other words, some image structures can be similar just in a portion of the time-scale plane, while they do not necessarily are the same in the whole plane.

The atomic approximation is a 1D model, then it can be independently applied to image rows and columns. Nonethe-

Table 1 $512 \times 512 \times 8$ bits Lena, Barbara, Peppers and Boats images. Comparisons (PSNR) between *ANL-means* and the denoising approaches in [5, 6, 12, 13, 28, 31, 32, 34], using different noise standard deviations

Image	Method	Noise StD σ					
		5	10	15	20	25	50
Lena	*ANL-means*	***39.00***	***36.33***	***34.89***	***33.65***	***32.45***	***28.97***
	Atomic model [34]	38.81	35.97	34.26	33.16	31.89	28.59
	BM-3D DFT [6]	38.72	35.93	34.27	33.05	32.08	28.86
	Bayes. NL means [13]	37.98	35.25	33.68	32.63	31.55	27.51
	NL means-W non iter [32]	–	35.55	–	–	31.70	28.26
	NL means-W with iter [32]	–	35.53	–	–	31.74	28.64
	PCA based NL means [28]	–	34.77	–	–	31.31	28.42
	SAWT [5]	–	35.10	33.37	32.08	31.07	–
	SURE-NLM [31]	–	35.1	33.2	–	30.7	–
	FSP-2LA [12]	38.55	35.66	33.96	32.71	31.72	28.61
Barbara	*ANL-means*	***38.45***	*34.96*	*32.99*	*31.51*	*30.50*	*26.80*
	Atomic model	38.03	34.38	32.08	30.80	29.50	25.29
	BM-3D DFT	38.31	34.98	**33.11**	**31.78**	**30.72**	**27.17**
	Bayes. NL means	36.93	33.82	32.21	30.88	29.77	24.91
	NL means-W non iter	–	**35.01**	–	–	30.56	26.57
	NL means-W with iter	–	34.91	–	–	30.59	26.99
	PCA based NL means	–	33.15	–	–	28.83	25.72
	SAWT	–	33.39	31.10	29.50	28.30	–
	SURE-NLM	–	33.2	31.0	–	28.1	–
	FSP-2LA	38.19	34.60	32.49	30.94	29.76	26.04
Peppers	*ANL-means*	*37.94*	*34.65*	***32.91***	***31.30***	*30.11*	***26.90***
	Atomic model	37.78	34.22	32.43	30.86	29.60	26.35
	BM-3D DFT	**38.12**	**34.68**	32.70	31.29	**30.16**	26.41
	Bayes. NL means	37.13	33.87	32.06	30.75	29.77	23.84
	PCA based NL means	–	33.76	–	–	29.37	26.10
	FSP-2LA	37.72	34.24	32.18	30.67	29.50	26.11
Boats	*ANL-means*	***37.90***	***34.45***	***32.87***	***31.50***	***30.10***	***26.95***
	Atomic model	37.68	34.12	32.40	30.96	29.70	26.65
	BM-3D DFT	37.28	33.92	32.14	30.88	29.91	26.64
	Bayes. NL means	36.14	33.09	31.44	30.12	29.20	25.93
	PCA based NL means	–	32.43	–	–	28.95	26.16
	FSP-2LA	37.27	33.90	32.05	30.68	29.57	26.61

less, the similarity measure in (10) gives it the second dimension, since it is able to detect edges as connected curves composed of neighboring and similar time-scale atoms. An example is shown in Fig. 6 where atoms along two edge curves of Lena image have been considered. Pixels belonging to the same edge exhibit the same atoms distribution in the time-scale plane. Hence, each of them contributes with a unitary weight in the denoising of the others without introducing constraints on image smoothness or edge shape, as in [20–22].

2.3 Parameters Estimation

To estimate atoms locations t_k and slopes $\alpha_{k,\overline{s}_1}$ at a fixed scale $\overline{s}_1$, a matching pursuit strategy is adopted [3] using (4): the atom with the greatest energy that best fits the data $w(u, s_1)$ in the cone of influence of an extremum point at t_k is selected at each step. Then, it is subtracted from the analyzed data and the algorithm is repeated on the residual signal. It is worth observing that the dictionary of the matching pursuit algorithm is just composed of the basic atoms in (3). On the contrary, by defining $\alpha_{k,\overline{s}} = \alpha_{k,s_1}\overline{s}^{\gamma_k - 1}$ the decays

Table 2 $512 \times 512 \times 8$ bits Lena and Barbara images. Comparisons (PSNR) between *ANL-means* and the denoising approaches in [10, 35–37], using different noise standard deviations

Image	Method	Noise StD σ					
		5	10	15	20	25	50
Lena	*ANL-means*	**39.00**	**36.33**	**34.89**	**33.65**	**32.45**	**28.97**
	BLS-GSM [35]	38.49	35.61	33.90	32.66	31.69	28.61
	SA-DCT [36]	38.54	35.58	33.86	32.62	31.66	28.60
	Learned Dict. [10]	38.60	35.47	33.70	32.38	–	27.79
	SURE [37]	38.29	35.08	–	32.00	–	28.03
Barbara	*ANL-means*	**38.45**	**34.96**	**32.99**	**31.51**	**30.50**	**26.80**
	BLS-GSM	37.79	34.03	31.86	30.32	29.13	25.48
	SA-DCT	37.47	33.48	31.37	30.00	28.95	25.44
	Learned Dict.	38.08	34.42	32.37	30.83	29.60	25.47
	SURE	37.69	33.90	31.71	30.16	28.96	25.32

γ_ks are derived by comparing the slopes of corresponding atoms at two (or more) successive scales as follows

$$\gamma_k = \frac{1}{|I_{\overline{s}}|} \sum_{s \in I_{\overline{s}}} \left(\log_s \frac{\alpha_{k,s}}{\alpha_{k,s_1}} \right) + 1, \tag{14}$$

where $I_{\overline{s}} = [s_1, s_f]$ is a proper scale interval that is small enough to preserve atom location.

To predict wavelet details at dyadic scales from not dyadic ones the same similarity criterion of previous section can be used: the atomic approximation at a fixed scale of a generic signal is "similar" to the wavelet transform computed at the same scale of a piecewise linear signal.

In other words, if $\alpha_{k,\overline{s}} = \alpha_{k,s_1} \overline{s}^{\gamma_k - 1}$ are atoms slopes at a fixed scale $\overline{s}$, then $w(u, \overline{s}) = \sum_k \alpha_{k,\overline{s}} F(t_k, u, \overline{s})$ is the wavelet transform at scale $\overline{s}$ of a piecewise linear signal $p(t)$ defined as follows

$$p(t) = \sum_{k=1}^{N} \left(\left(\sum_{h=1}^{k} \alpha_{h,\overline{s}} \right) (t - t_{h-1}) + \beta_h \right) \chi_{[t_{k-1}, t_k]}(t), \tag{15}$$

where $\beta_h = \sum_{h=1}^{k-1} \alpha_{h,\overline{s}} (t_k - t_h)$, $\chi_{[.](t)}$ is the unitary function in the range $[.]$, while $\alpha_{h,\overline{s}} = \alpha_h \overline{s}^{\gamma_h - 1}$. The equality in (15) can be derived if each basic atom $F(t_k, u, s)$ is written as the wavelet transform of ramp signals, whose equation is $r_k(t) = (t - t_k)\chi_{[t_k, +\infty]}$—see [34] for details.

In order to make the algorithm more robust to noise and to bad parameters estimation, in a dyadic wavelet decomposition, each dyadic band $\overline{s} = 2^j$, $j = 1, 2, \ldots$, can be independently processed using the scale interval $I_{2^j} = [2^j - \Delta s, 2^j + \Delta s]$, where Δs can be set equal to 2^{j-1}.

2.4 *ANL-Means* Algorithm

– Perform the undecimated discrete wavelet transform (UDWT) of the noisy signal g up to Jth scale level. For each detail band d_j at scale level j:

1. Estimate the slopes $\alpha_{k,s}$ and the relative locations t_k of the atomic approximation at scales $s \in I_{2^j} = [2^j - \Delta s, 2^j + \Delta s]$ of the continuous wavelet transform of the considered signal, using the matching pursuit algorithm and (4).
2. Estimate the corresponding γ_ks using (14) with $s_1 = 2^j - \Delta s$.
3. Eliminate atoms having $\gamma_k < 0$.
4. Estimate the cleaned detail band $\tilde{d}_j$ at scale $s = 2^j$ as follows:
   (i) for each estimated atom, compute the weights ω_{kl}, as in (13) and correct the slopes α_{k,s_1}s using (12) with $s_1 = \overline{s} - \Delta s$. Let $\overline{\alpha}_{k,s_1}$ be the corrected slopes;
   (ii) set the piecewise linear signal $p(t)$, as in (15), where $\overline{\alpha}_{k,2^j} = \overline{\alpha}_{k,2^j - \Delta s} 2^{j(\gamma_k - 1)}$;
   (iii) compute its dyadic undecimated discrete wavelet transform $\tilde{d}_j$ at the jth scale level.

– Invert the UDWT using the cleaned detail bands $\{\tilde{d}_j\}_{j=1,2,\ldots,J}$ to obtain the de-noised signal f_d.

The algorithm works on rows and columns respectively of the vertical and horizontal sub-bands of image 2D Wavelet Transform (2D-WT) [34], that is a redundant transform. In fact, the atomic approximation is not robust to decimation since it can cause the loss of extremum points from a dyadic scale to another and the change of atoms shape.

Fig. 7 *From left to right:* Original, noisy and recovered test images using *ANL-means*. *From top to bottom:* Lena corrupted with Gaussian noise with standard deviation $\sigma = 25$, Barbara ($\sigma = 20$), Peppers ($\sigma = 50$) and Boats with ($\sigma = 55$)

3 Experimental Results and Concluding Remarks

Extensive tests have been performed using several images. Some results achieved on commonly used $512 \times 512 \times 8$ bits test images Lena, Barbara, Peppers and Boats, will be presented. For all of them, the 2D-WT [33] has been computed up to the 4th scale level using the 3/9 spline biorthogonal wavelet.

Table 1 shows PSNR values of the denoised images that have been estimated from the corrupted ones using zero mean white Gaussian noise with different standard deviations $\sigma = 5, 10, 15, 20, 25, 50$. *ANL-means* has been performed over 15 different noise realizations for each standard deviation and the resulting PSNRs have been averaged over these runs. PSNR values have been compared with the most recent and powerful de-noising approaches that follow NL-means philosophy. In particular we have selected: the Bayesian non local means and variable window sizes in [13], the image denoising with block matching and 3D filtering BM-3D FFT in [6], the improved NL-means with iterations [32], the PCA based NL means [28], the spatially adaptive wavelet thresholding with context modelling SAWT in [5] (that is very close, in spirit, to NL-means), the SURE based NL-means [31] and the two level adaptation Gaussian mix-

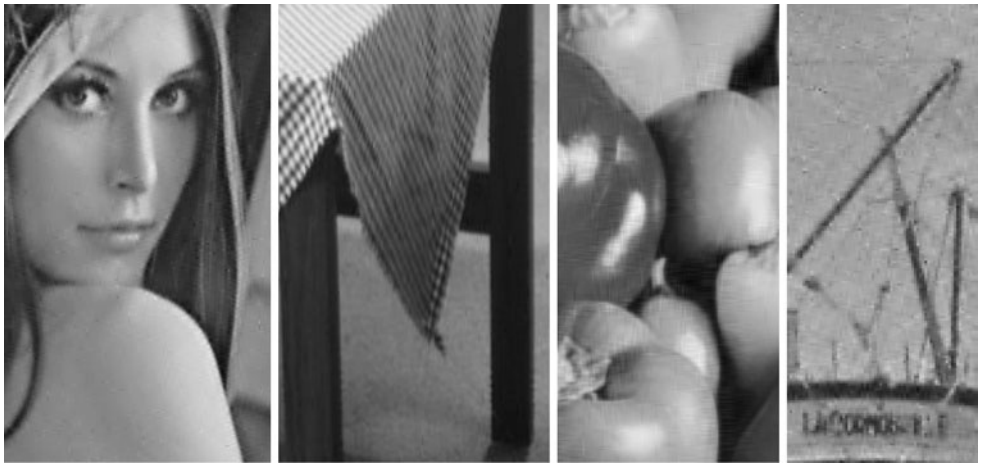


Fig. 8 Zoom of the denoised Lena, Barbara, Peppers and Boats images in Fig. 7

ture estimator FSP-2LA presented in [12]. PSNR values in Table 1 are taken from the original works along with their numerical precision. All selected approaches catch similarities in the image except for [12], that uses a coarser adaptation level by defining a larger neighborhood for estimating the local signal statistics at each subband of an overcomplete pyramid. *ANL-means* performance is comparable and often over-exceeds the most powerful NL-means based image de-noising schemes. Table 2 shows that *ANL-means* is also competitive with some of the most powerful denoisers, such as BLS-GSM [35], SA-DCT [36], SURE [37] and Learned Dictionaries [10]. In order to better appreciate the achieved results, Fig. 7 shows some denoised test images while a zoom around their sharp transitions is presented in Fig. 8. *ANL-means* is able to recover edges without introducing annoying ringing effects. The modeling of the multiscale behavior of significant points allows to discriminate between noise and interfering objects, while the modified matching pursuit preserves the correlation between adjacent coefficients. The modeling of the time-scale behavior of higher order singularities also avoids the over-smoothing of smooth regions and guarantees textures recovering. In fact, the content of each dyadic scale is derived from finer scales (atoms trajectories) allowing high frequencies preservation. Finally, since image similarities are measured from groups of coefficients, the best block size in which to evaluate similarity measures is atoms support at the considered resolution. Moreover, it compensates the one dimensional nature of the algorithm that does not make any geometrical assumption on the image—it independently processes rows and columns in the direction of the high pass filtering.

In order to emphasize the contribution of NL-means, Table 1 also includes the denoising results provided by the proposed approach without using self-similarities, that is presented in [34] (Atomic model). The gain goes from .09 db to 1.6 db over the considered set of images, confirming the potential of the Non Local means philosophy. It also confirms the fact that if similarities are not found, denoising still occurs since it is already embedded in the estimation of atoms slopes. Moreover, any user's interaction is required.

The main computational effort of *ANL-means* is required for the estimation of atoms' context, that involves independent least squares approximations over a fixed number of scales. For an $M \times M$ image, the complexity is $O(2MN|I_{\bar{s}}|)$, where N is the number of atoms and $|I_{\bar{s}}|$ is the employed number of scales. In particular, differently from classical NL-means, *ANL-means* does not use all coefficients in the computation of the weights in (13), but just the estimated atoms. It turns out that $O(N\frac{N-1}{2})$ comparisons are required for a fixed band. The CPU time on a Pentium IV processor for a 512×512 image is about 65 seconds using a non optimized Matlab code.

The proposed model is not adequate for non Gaussian and uncorrelated noise. This will be one of the main topics of our future research as well as a further optimization to speed up the algorithm. With regard to the employed function for weights computation, a deeper investigation would be required. As described in [32], different measures have been considered, but the de-noising results do not significantly change. In the future, it will be probably more interesting to introduce some metrics accounting for the human visual system, as done in [38], to really improve the visual quality of the recovered image.

Acknowledgements Authors would like to thank the anonymous reviewers for their suggestions and comments that allowed to improve the presentation of the work.

References

1. Azzabou, N., Paragios, N., Guichard, F.: Image denoising based on adapted dictionary computation. In: Proc. of ICIP 2007, vol. III, pp. 109–112 (2007)
2. Balster, E.J., Zheng, Y.F., Ewing, R.L.: Feature-based wavelet shrinkage algorithm for image denoising. IEEE Trans. Image Process. **14**(12), 2024–2039 (2005)

3. Bruni, V., Vitulano, D.: Wavelet based signal denoising via simple singularities approximations, signal processing. Elsevier Sci. **86**, 859–876 (2006)
4. Buades, A., Coll, B., Morel, J.M.: A review of image denoising algorithms with a new one. Multiscale Model. Simul. **4**(2), 490–530 (2005)
5. Chang, S.G., Yu, B., Vetterli, M.: Spatially adaptive thresholding with context modeling for image denoising. IEEE Trans. Image Process. **9**(1), 1522–1530 (2000)
6. Dabov, K., Foi, A., Katkovnik, V., Egiazarian, K.: BM3D image denoising with shape-adaptive principal component analysis. In: SPARS'09—Signal Processing with Adaptive Sparse Structured Representations, Saint-Malo, France, Apr. (2009)
7. Donoho, D.L.: Denoising by soft thresholding. IEEE Trans. Inf. Theory **41**(3), 613–627 (1995)
8. Dragotti, P.L., Vetterli, M.: Wavelet footprints: theory, algorithms and applications. IEEE Trans. Signal Process. **51**(5), 1306–1323 (2003)
9. Easley, G.R., Labate, D., Colonna, F.: Shearlet-based total variation diffusion for denoising. IEEE Trans. Image Process. **18**(2), 260–268 (2009)
10. Elad, M., Aharon, M.: Image denoising via learned dictionaries and sparse representation. In: Proceedings of IEEE CVPR'06 (2006)
11. Foi, A., Katkovnik, V., Egiazarian, K.: Pointwise shape adaptive DCT for high quality denoising and deblocking of grayscale and color images. In: Proc. of SPIE—IS & T Electronic Imaging, vol. 6064 (2006)
12. Guerrero-Colon, J.A., Mancera, L., Portilla, J.: Image restoration using space-variant Gaussian scale mixtures in overcomplete pyramids. IEEE Trans. Image Process. **17**(1), 27–41 (2008)
13. Kervrann, C., Boulanger, J., Coupè, P.: Bayesian non local mean filter, image redundancy and adaptive dictionaries for noise removal. In: Lecture Notes in Computer Science, Graz, Austria, vol. 4485, pp. 520–532. Springer, Berlin (2007)
14. Mallat, S., Hwang, W.L.: Singularity detection and processing with wavelets. IEEE Trans. Inf. Theory **38**(2), 617–643 (1992)
15. Perona, P., Malik, J.: Scale-space and edge detection using anisotropic diffusion. IEEE Trans. Pattern Anal. Mach. Intell. **12**, 629–639 (1990)
16. Pizurica, A., Philips, W., Lemanhieu, I., Acheroy, M.: A joint inter- and intrascale statistical model for Bayesian wavelet based image denoising. IEEE Trans. Image Process. **11**(5), 545–557 (2002)
17. Sendur, L., Selesnick, I.W.: Bivariate shrinkage with local variance estimation. IEEE Signal Process. Lett. **9**(12), 670–684 (2002)
18. Wei, J.: Lebesgue Anisotropic Image Denoising. Wiley, New York (2005)
19. Teboul, S., Blanc-Feraud, L., Aubert, G., Barlaud, M.: Variational approach for edge-preserving regularization using coupled PDE's. IEEE Trans. Image Process. **7**(3), 387–397 (1998)
20. Do, M.N., Vetterli, M.: The contourlet transform: an efficient directional multiresolution image representation. IEEE Trans. Image Process. **14**(12), 2091–2106 (2005)
21. Le Pennec, E., Mallat, S.: Sparse geometric image representations with bandelets. IEEE Trans. Image Process. **14**(4), 423–438 (2005)
22. Starck, J.L., Candes, E.J., Donoho, D.L.: The curvelet transform for image denoising. IEEE Trans. Image Process. **11**(6), 670–684 (2002)
23. Bilcu, R.C., Vehvilainem, M.: Fast non local means for image denoising. In: Proc. of SPIE Digital Photography III, vol. 6502(1) (2007)
24. Dauwe, A., Goossens, B., Luong, H., Philips, W.: A fast non local image denoising algorithm. In: Proc. of SPIE Electronic Imaging, San Jose, USA, Jan. 2008, vol. 6812 (2008)
25. Coupé, P., Hellier, P., Prima, S., Kervrann, C., Barillot, C.: 3D wavelet subbands mixing for image denoising. Int. J. Biomed. Imaging (2008). doi:10.1155/2008/590183
26. Wang, J., Guo, Y., Ying, Y., Liu, Y., Peng, Q.: Fast non-local algorithm for image denoising. In: Proc. of IEEE ICIP, Atlanta, CA, Oct. 2006, pp. 1429–1432 (2006)
27. Mahmoudi, M., Sapiro, G.: Fast image and video denoising via non local means of similar neighborhoods. IEEE Signal Process. Lett. **12**(12), 839–842 (2005)
28. Tasdizen, T.: Principal neighborhood dictionaries for non-local means image denoising. IEEE Trans. Image Process. **18**(12), 2649–2660 (2009)
29. Chatterjee, P., Milanfar, P.: A generalization of non-local means via kernel regression. In: Proc. of IS&T/SPIE Conf. on Computational Imaging VI, San Jose, USA, Jan. 2008, vol. 6814 (2008)
30. Brox, T., Cremers, D.: Iterated nonlocal means for texture restoration. Lect. Notes Comput. Sci. **4485**, 13–24 (2007)
31. Van De Ville, D., Kocher, M.: SURE-based non-local means. IEEE Signal Process. Lett. **16**(11), 973–976 (2009)
32. Goossens, B., Luong, H., Pizurica, A., Philips, W.: An improved non-local denoising algorithm. In: International Workshop on Local and Non-Local Approximation in Image Processing (LNLA2008), Lausanne, Switzerland, Aug. 2008, pp. 25–29 (2008)
33. Mallat, S.: A Wavelet Tour of Signal Processing. Academic Press, San Diego (1998)
34. Bruni, V., Piccoli, B., Vitulano, D.: Fast computation method for time scale signal denoising. In: Signal, Image and Video Processing, vol. 3(1). Springer, London (2009)
35. Portilla, J., Strela, V., Wainwright, M., Simoncelli, E.: Image denoising using scale mixtures of Gaussians in the wavelet domain. IEEE Trans. Image Process. **12**(11), 1338–1351 (2003)
36. Foi, A., Katkovnik, V., Egiazarian, K.: Pointwise shape adaptive DCT for high quality denoising and deblocking of grayscale and color images. IEEE Trans. Image Process. **16**(5), 1395–1411 (2007)
37. Luisier, F., Blu, T., Unser, M.: A new SURE approach to image denoising: interscale orthonormal wavelet thresholding. IEEE Trans. Image Process. **16**(3), 593–606 (2007)
38. Wong, A., Fieguth, P., Clausi, D.: A perceptually adaptive approach to image denoising using anisotropic non-local means. In: Proc. of IEEE ICIP, San Diego, CA, Oct. 2008, pp. 537–540 (2008)

Vittoria Bruni received the degree in Mathematics from the University of Rome "La Sapienza" in 2001. In January 2006 she received the Ph.D. degree in applied mathematics from the same University, Department of Mathematical Methods and Models for Applied Sciences. Since 2010 she is researcher at the Department of SBAI—Faculty of Engineering of the same University. Her research interests include image compression, denoising and restoration, wavelets theory and applications, pattern recognition with applications in the Field of Cultural Heritage. She is co-author of several scientific articles published on international journals and in the proceedings of international conferences. She is also referee for various international Journals. Since September 2001 she collaborates with the Institute for the Application of Calculus (IAC) of National Research Council in Rome within national and international projects.

Domenico Vitulano received the physics degree from the University of Naples "Federico II" in 1993 and M.Sc. degree (summa cum laude) in information and communication technology from IIASS "R.R. Caianiello" Institute of Vietri sul Mare, Salerno, in 1997. Since 1995 he is at Institute for the Application of Calculus of National Council of Research in Rome and since 2001 he is permanent researcher. His research interests include pattern recognition, image and data compression, indexing and computer vision. He has been involved in various national and international Research Projects concerning Cultural Heritage. He is currently head of the CNR research line (RSTL) entitled 'Multiscale Analysis for Complicated Shapes Recognition' (DG.RSTL.004.009) and of the CNR research line entitled 'Analysis and Sinthesys of Etherogenous Data for a Computer-Aided Monitoring of Cultural Heritage' (PC.P03.008). He is currently lecturer at University of Rome "Tor Vergata", Department of Mathematics, degree in "Mathematical Processing of Images and Signals". He is author and co-author of more than fifty scientific papers on journals, tutorials, monographs and proceedings of workshops. He has been referee for various international conferences and Journals.