# Anisotropic Wavelet Thresholding For Bayesian Image Denoising

Giovanni Jacovitti*, Alessandro Neri**

*INFOCOM Department, University of Rome "La Sapienza", Rome, ITALY
** Department of Electronic Engineering, University of ROMA TRE, Rome, ITALY
(e-mail: jacovitti@infocom.uniroma1.it, neri@ele.uniroma3.it).

***Abstract*** - A new technique for noise suppression in images by wavelet thresholding is presented. The technique focuses on perceptual relevant features, using a multiscale edge oriented wavelet representation. A orientation-dependent zero-memory non-linear soft thresholding rule is defined in the framework of Bayesian MMSE estimation.

***Keywords:*** Denoising, wavelet thresholding, shrinking, Bayesian estimation.

## I. INTRODUCTION

Traditional signal and image denoising is linearly performed using Wiener filtering. More recently, non linear techniques applied to wavelet based representations have been introduced. Basically, these techniques consist of tresholding the wavelet coefficients, setting them to zero if they are at noise level, and retaining them if they represent significant structures. This technique, is also called "wavelet shrinkage". Since the pioneering paper of Donoho [1] much work has been done for setting and optimizing thresholds under different criteria and using adaptive strategies (see for instance [2], [3]). These approaches are based on general wavelet image representations, such as Daubechies finite support wavelets. In the recent past, schemes taking also into account statistical dependencies between wavelet coefficients have been devised.

In this contribution, a new approach based on an *edge oriented* wavelet representation is presented. This choice has two advantages. From a perceptive viewpoint, it focuses on the most relevant visual image features. From the statistical processing viewpoint, it applies to coefficients more decorrelated in space.

To this purpose, images are represented in a multiscale edge domain. Stemming from [4], where pairs of wavelet defined by smoothed vertical and horizontal derivatives were employed, a complex edge oriented wavelet is defined. So, in the transformed wavelet domain edges are represented by complex numbers with magnitude proportional to their "strength" and phase equal to their orientation angle. The use of this complex representation simplifies algebra and notations.

Signal and noise edges are modelled as 2D *Gaussian mixtures*. This flexible approach allows us to model different signals, and non Gaussian noises as well. The Bayesian MMSE estimate of the uncorrupted image is then calculated in a general form, leading to an estimator constituted by an anisotropic, zero memory non-linearity.

## II. THE EDGE WAVELET

We refer to a multiscale feature decomposition based on CHFs (Circular Harmonic Filters), which are complex, polar separable filters defined by a Point Spread Function (PSF) with a polar representation of the kind

$$h^{(k)}(r,q) = v^{(k)}(r)\, e^{-jkq}, \qquad k = 0,1,2,... \quad (1)$$

where $r$ and $q$ are polar coordinates, $k$ is the order of the CHF and $v^{(k)}(\cdot)$ is the radial profile [5]. A first-order CHF acts as a differential operator suited for edge extraction. Its output is a complex image whose magnitude is proportional to edge strength, and phase is equal to edge orientation.

Let us now consider the specific family of CHFs with Gauss-Laguerre radial profiles. It is shown that each element of this family defines a dyadic Circular Harmonic Wavelet (CHW) suited for multiscale representation [5]. In particular, let us refer to a mother wavelet constituted by the following first-order member of the CHF *Gauss-Laguerre* family:

$$h^{(1)}(r,q) = L_0^{(1)}(r,q) = 2p^{1/2}\,(a \quad r)\;\; e^{-p \;\; (ar)^2}\, e^{jq}\,. \quad (2)$$

For notational convenience, let us denote with $h^{(1)}(n,m)$ the mother wavelet in discrete Cartesian coordinates, and with $h_s(n,m)$ the function obtained by dilation of $h^{(1)}(n,m)$ by a scaling factor $s$, i.e.

$$h_s(n,m) = \frac{1}{s} h^{(1)}\left(\frac{n}{s}, \frac{m}{s}\right). \quad (3)$$

Let $\{s_1, s_2, ..., s_L\}$ be a finite sequence of scale factors corresponding to increasing resolutions. For a given image $y(n,m)$, the set of the complex images $\{z_{s_k}(n,m) = y(n,m) * h_{s_k}(n,m),\ k = 1,2,...,L\}$, along with a coarse approximation at low resolution $z_{s_0}(n,m) = y(n,m) * h^{(0)}(n,m)$, obtained with a low pass zero order CHF filter $h^{(0)}(n,m)$, constitute its wavelet representation in the multiscale complex edge feature domain. In fact,

$$y(n,m) = \sum_{k=0}^{L} z_{s_k}(n,m) * g_{s_k}(n,m), \quad (4)$$

where the reconstruction filters $g_{s_k}(n,m)$ are defined in terms of their Fourier transform as follows

$$G_{s_h}(w_1,w_2) = \frac{H^*_{s_h}(w_1,w_2)}{\sum_{k=0}^{L} \left| H_{s_k}(w_1,w_2) \right|^2}, \quad k = 1,...,L, \quad (5)$$

where $H_{s_k}(w_1,w_2)$ are the frequency responses of the filters $h_{s_k}(n,m)$ and $H_{s_0}(w_1,w_2) = H^{(0)}(w_1,w_2)$ is the frequency response of the filter $h^{(0)}(n,m)$.

## III. THE BAYESIAN RESTORATION IN THE EDGE FEATURE DOMAIN

Introducing a shorthand notation, let us consider an image $\mathbf{Y}=\{y(n,m)\}$, and its observed version $\mathbf{Y}^M =\{y^M(n,m)\}$ corrupted by an additive independent observation noise $\mathbf{W}=\{w(n,m)\}$. A Bayesian estimate $\tilde{\mathbf{Y}}$ of $\mathbf{Y}$, given $\mathbf{Y}^M$, is obtained by the minimization of the associated absolute risk. Let $\mathbf{Z}^M_{s_k} = \left\{ z^M_{s_k}(n,m) \right\}$ be the complex edge image at resolution $s_k$ of $\mathbf{Y}^M$. Based on (4), we want to recover $\mathbf{Y}$ from the MMSE estimates of the complex edge images at different resolutions, respectively constituted by the a posteriori expectation $\tilde{\mathbf{Z}}_{s_k} = \left\{ \tilde{z}_{s_k}(n,m) \right\}$ of $\mathbf{Z}_{s_k}$, given $\mathbf{Z}^M_{s_k}$. Looking for a suboptimum solution, let us now ignore residual spatial correlation. Thus, *for each scale* we evaluate the conditional expectation $\tilde{z}_{s_k}(n,m)$ of $z_{s_k}(n,m)$ given $z^M_{s_k}(n,m)$ at site $(n,m)$ only. Let $\mathbf{z}_{s_k}(n,m) = \left[ z_{R_{s_k}}(n,m) \quad z_{I_{s_k}}(n,m) \right]^T$ be the 2D vector corresponding to complex numbers $z_{s_k}(n,m) = z_{R_{sk}}(n,m) + jz_{I_{sk}}(n,m)$. Then, we describe the marginal distribution of $\mathbf{z}_{s_k}(n,m)$ with a rather general model constituted by a bivariate zero mean Gaussian mixture, i.e. a weighted sum of bivariate Gaussian distributions:

$$p_Z\left[\mathbf{z}_{s_k}(n,m)\right] = \sum_{i=1}^{K} l_i N_2\left[\mathbf{z}_{s_k}(n,m), \mathbf{0}, \mathbf{R}_{Z_i}(n,m)\right], \quad (6)$$

where $N_2[\mathbf{z}; m, \mathbf{R}]$ denotes the Gaussian probability density function of a 2D random variate $\mathbf{z}=[z_R\ z_I]^T$ with expectation $m=[m_R\ m_I]^T$ and covariance matrix $\mathbf{R}$. Let $\Delta z_{s_k}(n,m) = w(n,m) * h_{s_k}(n,m)$ be the coefficients of the CHW transform of the observation noise $\mathbf{W}$ modeled again as a zero mean bivariate Gaussian mixture with mixing parameters $b_j$, namely,

$$p_{\Delta Z}\left[\Delta\mathbf{z}_{s_k}(n,m)\right] = \sum_{j=1}^{M} b_j N_2\left[\Delta\mathbf{z}_{s_k}(n,m), \mathbf{0}, \mathbf{R}_{N_j}(n,m)\right] \quad (7)$$

Application of Bayes rule yields the following expression for the conditional expectation $\hat{\mathbf{z}}_{s_k}(n,m)$ of $\mathbf{z}_{s_k}(n,m)$, given $\mathbf{z}^M_{s_k}(n,m)$:

$$\hat{\mathbf{z}}_{s_k}(n,m) = E_{\mathbf{Z}_{s_k}/\mathbf{Z}^M_{s_k}}\left\{ \mathbf{z}_{s_k}(n,m) / \mathbf{z}^M_{s_k}(n,m) \right\}$$
$$= \sum_j \sum_i w_{ij}\left[\mathbf{z}^M_{s_k}(n,m)\right] \mathbf{R}_{Z_i}(n,m)$$
$$\times \left(\mathbf{R}_{N_j}(n,m) + \mathbf{R}_{Z_i}(n,m)\right)^{-1} \mathbf{z}^M_{s_k}(n,m), \quad (8)$$

where

$$w_{ij}\left[\mathbf{z}^M_{s_k}(n,m)\right] = \frac{b_j l_i N_2\left[\mathbf{z}^M_{s_k}(n,m), \mathbf{0}, \mathbf{R}_{N_j}(n,m) + \mathbf{R}_{Z_i}(n,m)\right]}{\sum_j \sum_i b_j l_i N_2\left[\mathbf{z}^M_{s_k}(n,m), \mathbf{0}, \mathbf{R}_{N_j}(n,m) + \mathbf{R}_{Z_i}(n,m)\right]}. \quad (9)$$

Eq. (8) says that in general, for signal and noise with bivariate Gaussian mixture distributions, the MMSE estimator is a *non-linear combination of conditional linear estimators, with gains*

$$\mathbf{R}_{Z_i}(n,m)\left(\mathbf{R}_{N_j}(n,m) + \mathbf{R}_{Z_i}(n,m)\right)^{-1}$$

*each matched to a pair (i,j) of Gaussian submodels*. The weights $w_{ij}\left[\mathbf{z}^M_{s_k}(n,m)\right]$ are just the posterior probabilities of each submodel pair.

As illustrated in Fig.1, where the magnitude $\left|\hat{\mathbf{z}}_{s_k}(n,m)\right|$ of the estimated edge versus the observed (noisy) edge magnitude and orientation, for the reference case of isotropic signal in anisotropic noise (e.g. noise aligned along the vertical direction), is reported, the Bayesian Zero Memory Non Linear (ZNL) estimator attenuates small edges more or less deeply, depending on their direction, thus acting as an *anisotropic shrinking* function. Attenuation tends to zero for features orthogonal to the noise direction. For comparison, the conventional isotropic soft thresholding law is shown in Fig. 2.

To perform the Bayesian estimate, the actual parameters of the mixture distributions must be known. Usually, they are adaptively determined by moment matching over windows centered around the point (*n,m*). Starting from the MMSE estimates of the CHW coefficients the restored image is finally reconstructed with the reconstruction formula (4).

## IV. APPLICATION EXAMPLES

The method presented here is intended for image quality improvement in multimedia applications, for instance in coding preprocessing (as outlined in [3] and [4]) and in decoding post-processing. Here, let us first consider as an example the classical problem of cleaning images affected by Gaussian noise. This means that in the complex edge feature domain the marginal probability density function of $\Delta z(n,m)$ is simply modeled by (7) with $M=1$ and $\mathbf{R}_N(n,m) = s_N^2 \mathbf{I}$, where $\mathbf{I}$ is the identity matrix. We have used in this case a *two-term mixture*, and have locally tuned the *ZNL* estimator by moment matching in an 8x8 sliding window. In Figs.3 and 4 the noisy image and the restored one are shown. *Notice that textures are explicitly modeled as non Gaussian signals with preferred orientations.*

Let us consider now a new solution to the problem of restoring DCT block coded images affected by quantization artifacts. As well known, block DCT coders partition the original image $\mathbf{Y}$ into $N$x$N$ blocks, separately represented with the quantized coefficients of their DCTs. The decoded image $\mathbf{Y}^M$ is then reconstructed by applying the inverse DCT to each block. Blocking artifacts manifest themselves as spurious edges at block boundaries. Provided that quantization is small, blocking noise can be considered additive and uncorrelated with the image. Since first order HAF is a differential operator, the blocking artifacts appearing in $\mathbf{Y}^M$ are "focused" in $\mathbf{Z}^M$ on the block boundaries. In this case, the statistical model is assumed a priori. From statistical experiments conducted on different images, the distribution of $\Delta\mathbf{Z} = \mathbf{Z}^M - \mathbf{Z}$ is modeled by a two term Gaussian mixture with diagonal $\mathbf{R}_{N_j}(n,m)$ matrices, i.e.:

$$\mathbf{R}_{N_j}(n,m) = \begin{bmatrix} s_{H_j}^2(n,m) & 0 \\ 0 & s_{V_j}^2(n,m) \end{bmatrix}.$$

Considering noise localization, we set $\sigma_{H_j} >> \sigma_{V_j}$ ($\sigma_{H_j} << \sigma_{V_j}$) along horizontal (vertical) block boundaries, and $\sigma_{H_j} \cong \sigma_{V_j}$ for pixels interior to a block.

Let us refer to the image Lena in its JPEG coded version at 0.33 bit/pel (Fig. 5). The Bayesian non-adaptive de-blocking process gives the image displayed in Fig. 6.

## V. CONCLUSION

With respect to previous wavelet coefficient soft-thresholding techniques, the method described here presents additional features due to:

- application to most perceptually relevant image features
- use of flexible non Gaussian probabilistic models for both signal and noise, leading to anisotropic thresholding rules.
- intrinsic spatial decorrelation of the representation coefficients in the employed wavelet domain.

## REFERENCES


[1] D.L.Donoho, "De-noising by soft-tresholding", *IEEE Trans. on Information Theory*, vol. 41, pp. 613-627, May 1995.

[2] S.G. Chang, Bin Yu, M. Vetterly, "Adaptive Wavelet Tresholding for Image Denoising and Compression", *IEEE Trans. on Image Processing*, Vol.9, No pp. 1532-1546, Sept. 2000.

[3] M.K. Mihcak, I. Kozintsef, K Ramchandran, P. Moulin, "Low-Complexity Image Denoising based on Statistical Modeling of Wavelet Coefficients", *IEEE Signal processing Letters*, pp. 300-303, Vol.6, No 12, December 1999.

[4] S. Mallat, S. Zhong "Characterization of Signals from Multiscale Edges", *IEEE Trans. on P.A.M.I.*, Vol. 14, No7, pp. 710-732, July 1992

[5] G. Jacovitti, A. Neri, "Multiresolution circular harmonic decomposition", *IEEE Trans. on Signal Processing*, Vol. 48. No. 11, pp.3242-3247, November 2000.

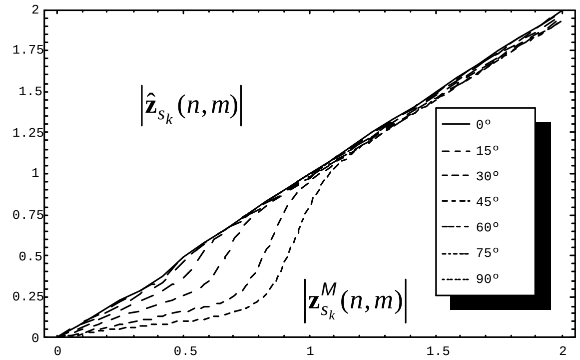


Fig. 1. $|\hat{\mathbf{z}}_{s_k}(n,m)|$ versus $\mathbf{z}^M_{s_k}(n,m)$ magnitude and orientation for noise aligned along vertical direction and isotropic signal ($\sigma_{H_1}$=0.02, $\sigma_{V_1}$=0.2, $\sigma_{H_2}$=0.05, $\sigma_{V_2}$=0.1, $\beta_1$=0.8, $\sigma_{Z_1}$=0.1, $\sigma_{Z_2}$=1, $\lambda_1$=0.99).

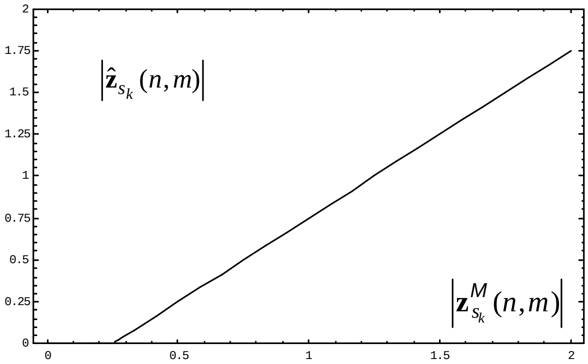


Fig. 2. $|\hat{\mathbf{z}}_{s_k}(n,m)|$ versus $|\mathbf{z}^M_{s_k}(n,m)|$ - conventional isotropic soft thresholding law.

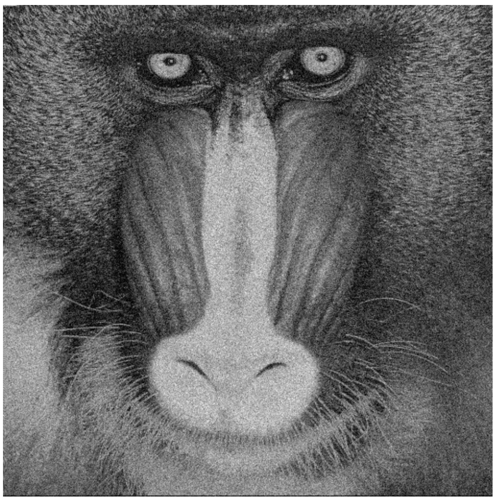

Fig.3 - “Baboon” image (512X512) with additive white noise

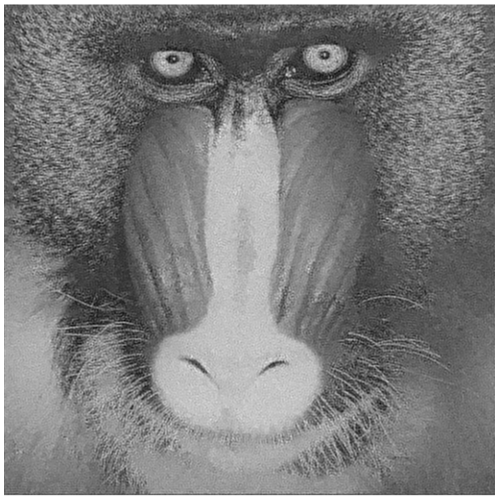

Fig.4 - Restored image.

Fig.5 – JPEG decoded image at 0.33 bits/pixel.

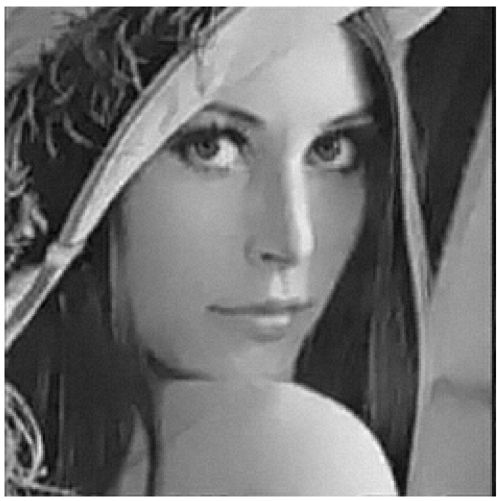

Fig.6 –Restored image.