4

A MAP-MRF Approach for Wavelet-Based Image Denoising

Alexandre L. M. Levada[1], Nelson D. A. Mascarenhas[2] and Alberto Tannús[3]
[1,2]*Federal University of São Carlos (UFSCar)*
[3]*University of São Paulo (USP)*
Brazil

1. Introduction

Image denoising is a required pre-processing step in several applications in image processing and pattern recognition, from simple image segmentation tasks to higher-level computer vision ones, as tracking and object detection for example. Therefore, estimating a signal that is degraded by noise has been of interest to a wide community of researchers. Basically, the goal of image denoising is to remove the noise as much as possible, while retaining important features, such as edges and fine details. Traditional denoising methods have been based on linear filtering, where the most usual choices were Wiener, convolutional finite impulse response (FIR) or infinitie impulse response (IIR) filters. Lately, a vast literature on non-linear filtering has emerged Barash (2002); Dong & Acton (2007); Elad (2002); Tomasi & Manduchi (1998); Zhang & Allebach (2008); Zhang & Gunturk (2008), especially those based on wavelets Chang et al. (2000); H. et al. (2009); Ji & Fermüller (2009); Nasri & Nezamabadi-pour (2009); Yoon & Vaidyanathan (2004) inspired by the remarkable works of Mallat (1989) and after Donoho (1995).

The basic wavelet denoising problem consists in, given an input noisy image, dividing all its wavelet coefficients into relevant (if greater than a critical value) or irrelevant (if less than a critical value) and then process the coefficients from each one of these groups by certain specific rules. Usually, in most denoising applications *soft* and *hard* thresholding are considered, in a way that filtering is performed by comparing each wavelet coefficient to a given threshold and supressing it if its magnitude is less than the threshold; otherwise, it is kept untouched (hard) or shrinked (soft). *Soft-thresholding* rule is generally preferred over *hard-thresholding* for several reasons. First, it has been shown that *soft-thresholding* has several interesting and desirable mathematical properties Donoho (1995), Donoho & Johnstone (1994). Second, in practice, the *soft-thresholding* method yields more visually pleasant images over *hard-thresholding* because the latter is discontinuous and generates abrupt artifacts in the recovered images, especially when the noise energy is significant. Last but not least, some results found in the literature Chang et al. (2000) conclude that the optimal *soft-thresholding* estimator yields a smaller estimation error than the optimal *hard-thresholding* estimator.

However, for some classes of signals and images, *hard-thresholding* results in superior estimates to that of *soft-thresholding*, despite some of its disadvantages Yoon & Vaidyanathan (2004). To tackle this problem, several hybrid thresholding functions have been proposed in the literature.

To test and evaluate our method, we built a series of experiments using both real Nuclear Magnetic Resonance (NMR) images and simulated data, considering several wavelet basis. The obtained results show the effectiveness of *GSAShrink*, indicating a clear improvement on the wavelet denoising performance in comparison to the traditional approaches. As in this chapter we are using a sub-optimal combinatorial optimization algorithm to approximate the optimal MAP solution, *GSAShrink* converges to a local maximum, making our method sensitive to different initializations. What at first could look like a disadvantage, actually revealed to be an interesting and promissing feature, mostly because we can incorporate other non-linear filtering techniques in a really straighforward way, by simply using them to generate better initial conditions for the algorithm. Results obtained by combining Bilateral Filtering and *GSAShrink* show that the MAP-MRF method under investigation is capable of suppressing the noise while preserving most relevant image details, avoiding the appearance of visible artifacts.
The remaining of the chapter is organized as follows. Section 2 describes the Discrete Wavelet Transform (DWT) in the context of digital signal processing, showing that, in practice, this transform can be implemented by a *Perfect Reconstruction Filter Bank* (PRFB), being completely characterized by a pair of *Quadrature Mirror Filters* (QMF), $h_0[]$, a low-pass filter and $g_1[]$, a high-pass filter. Section 3 briefly introduces the wavelet-based denoising problem, describing the proposed MAP-MRF solution, as well as the statistical modeling and threshold estimation, a crucial step in this kind of application. In Section 4 we briefly discuss the MRF Maximum Pseudo-Likelihood parameter estimation. The experimental setup and the obtained results are described in Section 5. Finally, Section 6 brings the our conclusions and final remarks.

2. The Wavelet transform

The basic tool for our MAP-MRF approach is the wavelet transform. Roughly speaking, in mathematical terms, the wavelet transform is an expansion that decomposes a given signal in a basis of orthogonal functions. In this sense, we can set a complete analogy with the Fourier Transform. While the Fourier Transform uses periodic, smooth and unlimited basis functions (i.e., sines and cosines), the wavelet transform uses non-periodic, non-smooth and finite support basis functions (i.e., Haar, Daubechies,...), allowing a much more meaningful representation through multi-resolution analysis, since it can capture a wide . In practice, the Discrete Wavelet Transform (DWT) can be implemented by a Perfect Reconstruction Filter Bank (PRFB), being completely characterized by a pair of Quadrature Mirror Filters (QMF) $h_0[]$, a low-pass filter, and $g_1[]$, the corresponding high-pass filter, known as analysis filters.

2.1 Perfect reconstruction filter banks (PRFB)

This section describes the Discrete Wavelet Transform from a digital signal processing perspective, by characterizing its underlying mathematical model by means of the Z-Transform. For an excellent review on wavelet theory and mathematical aspects of filter banks the reader is refered to Jensen & Cour-Harbo (2001); Strang & Nguyen (1997), from where most results described in this section were taken. A two-channel *perfect reconstruction □lter bank* (PRFB) consists of two parts: an analysis filter bank, responsible for the decomposition of the signal in wavelet sub-bands (DWT) and a synthesis filter bank, that reconstructs the signal by synthesizing these wavelet sub-bands Ji & Fermüller (2009). Figure 1 shows the block diagram of a two-channel PRFB, where $H_0(z)$ and $G_1(z)$ are the Z-transforms of the pair of analysis filters, $r_0[n]$ and $r_1[n]$ are the resulting signals after low-pass and high-pass filtering, respectively, $y_0[n]$ and $y_1[n]$ are the downsampled signals, $t_0[n]$ and $t_1[n]$

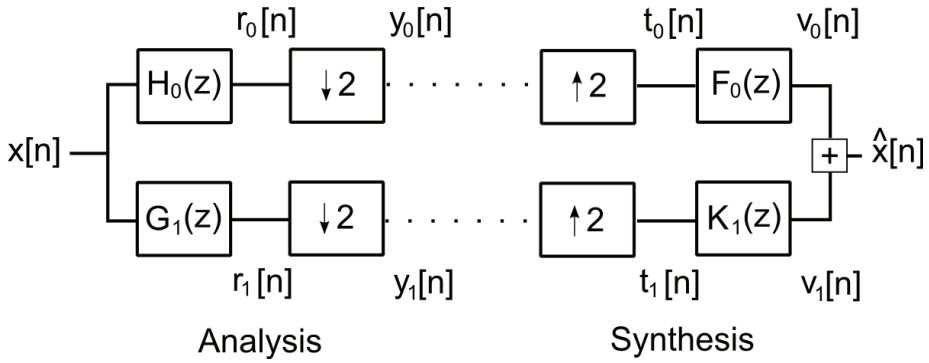


Fig. 1. Block diagram of a two-channel *Perfect Reconstruction Filter Bank*

are the upsampled signals obtained by placing zeros between each pair of samples, $F_0(z)$ and $K_1(z)$ are the Z-transforms of the pair of synthesis filters, and finally, $v_0[n]$ and $v_1[n]$ are interpolated signals that are combined to produce the reconstructed output $\hat{x}[n]$.
The basic assumption for perfect reconstruction is that the output $\hat{x}[n]$ has to be a delayed version of the input signal $x[n]$. Suppose that in the filter bank depicted in Figure 1, we have ℓ levels, each one causing a delay. Then, in mathematical terms, the condition for perfect reconstruction is:

$$\hat{x}[n] = x[n - \ell] \tag{1}$$

which means that the entire system can be replaced by a single transfer function. Equivalently, in the Z-domain we have:

$$\hat{X}(z) = z^{-\ell} X(z) \tag{2}$$

As the filter bank defines a linear time invariant (LTI) system and using the convolution theorem, we have:

$$R_0(z) = H_0(z)X(z) \tag{3}$$
$$R_1(z) = G_1(z)X(z) \tag{4}$$

and using the Z-transform property of decimation operators:

$$Y_0(z) = \frac{1}{2}\left\{R_0\left(z^{1/2}\right) + R_0\left(-z^{1/2}\right)\right\} \tag{5}$$
$$Y_1(z) = \frac{1}{2}\left\{R_1\left(z^{1/2}\right) + R_1\left(-z^{1/2}\right)\right\} \tag{6}$$

leading to the following relationship:

$$Y_0(z) = \frac{1}{2}\left\{H_0(z^{1/2})X(z^{1/2}) + H_0(-z^{1/2})X(-z^{1/2})\right\} \tag{7}$$

$$Y_1(z) = \frac{1}{2}\left\{G_1(z^{1/2})X(z^{1/2}) + G_1(-z^{1/2})X(-z^{1/2})\right\} \tag{8}$$

Since $H_0(z)$ and $G_1(z)$ are not ideal half-band filters, downsampling can introduce aliasing since we cannot reduce the interval between samples by half because we would be sampling below the Nyquist rate. To overcome this problem, conditions for alias cancellation must be enforced. According to the perfect reconstruction condition:

$$V_0(z) + V_1(z) = z^{-\ell}X(z) \tag{9}$$

Using the upsampling property of the Z-transform, we have the following expressions for $V_0(z)$ and $V_1(z)$:

$$V_0(z) = F_0(z)T_0(z) = F_0(z)Y_0(z^2) \tag{10}$$

$$V_1(z) = K_1(z)T_1(z) = K_1(z)Y_1(z^2) \tag{11}$$

which leads to:

$$V_0(z) = \frac{1}{2}F_0(z)\left\{H_0(z)X(z) + H_0(-z)X(-z)\right\} \tag{12}$$

$$V_1(z) = \frac{1}{2}K_1(z)\left\{G_1(z)X(z) + G_1(-z)X(-z)\right\} \tag{13}$$

Thus, grouping similar terms and enforcing the perfect reconstruction condition, we have the following equation that relates the input, analysis filters, synthesis filters and the output of the LTI system:

$$\begin{aligned}&\frac{1}{2}\left\{F_0(z)H_0(z) + K_1(z)G_1(z)\right\}X(z) + \\ &\frac{1}{2}\left\{F_0(z)H_0(-z) + K_1(z)G_1(-z)\right\}X(-z) = z^{-\ell}X(z)\end{aligned} \tag{14}$$

Therefore, a *perfect reconstruction filter bank* must satisfy the following conditions:

1. *Alias cancellation*

$$F_0(z)H_0(-z) + K_1(z)G_1(-z) = 0 \tag{15}$$

2. *Perfect Reconstruction (No distortion)*

$$F_0(z)H_0(z) + K_1(z)G_1(z) = 2z^{-\ell} \tag{16}$$

The first condition is trivially satisfied by defining the synthesis filters as:

$$F_0(z) = G_1(-z) \tag{17}$$
$$K_1(z) = -H_0(-z) \tag{18}$$

This condition implies that:

$$\begin{aligned} F_0(z) &= G_1(-z) \\ &= \sum_{-\infty}^{\infty} g_1[n](-z)^{-n} \\ &= \sum_{-\infty}^{\infty} (-1)^n g_1[n] z^{-n} \end{aligned} \tag{19}$$

and

$$\begin{aligned} K_1(z) &= -H_0(-z) \\ &= -\sum_{-\infty}^{\infty} h_0[n](-z)^{-n} \\ &= \sum_{-\infty}^{\infty} (-1)^{n+1} h_0[n] z^{-n} \end{aligned} \tag{20}$$

so that the synthesis filters coefficients are obtained directly from the analysis filters by a simple alternating signs rule:

$$\begin{aligned} f_0[n] &= (-1)^n g_1[n] \\ k_1[n] &= (-1)^{n+1} h_0[n] \end{aligned} \tag{21}$$

Defining $P_0(z) = F_0(z)H_0(z)$ and using equation (19) on (16) leads to:

$$P_0(z) - P_0(-z) = 2z^{-\ell} \tag{22}$$

where ℓ must be odd since the left hand side of (22) is an odd function, since all even terms cancel each other. Let $P(z) = z^{\ell} P_0(z)$. Then, $P(-z) = -z^{\ell} P_0(-z)$, since ℓ is odd. Rewriting equation (22) we finally have:

$$P(z) + P(-z) = 2 \tag{23}$$

showing that for perfect reconstruction the low-pass filter $P(z)$ requires all even powers to be zero, except the constant term. The design process starts with the specification of $P(z)$ and

then the factorization of $P_0(z)$ into $F_0(z)H_0(z)$. Finally, the alias cancellation condition is used to define $G_1(z)$ and $K_1(z)$. It has been shown that flattest $P(z)$ leads to the widely recognized Daubechies wavelet filter Daubechies (1988).
In this chapter, we consider the traditional 2-D separable DWT, also known as Square Wavelet Transform, that is based on consecutive one dimensional operations on columns and rows of the pixel matrix. The method first performs one step of the 1-D DWT on all rows, yielding a matrix where the left side contains down-sampled low-pass (h filter) coefficients of each row, and the right contains the high-pass (g filter) coefficients. Next, we apply one step to all columns, resulting in four wavelet sub-bands: LL (which is known as approximation signal), LH, HL and HH. A multilevel decomposition scheme can be generated in a straghtforward way, always expanding the approximation signal.
The analysis of a signal or image wavelet coefficients suggests that small coefficients are dominated by noise, while coefficients with a large absolute value carry more signal information. Thus, supressing or smoothing the smallest, noisy coefficients and applying the Inverse Wavelet Transform (IDWT) lead to a reconstruction with the essential signal or image characteristics, removing the noise. More precisely, this idea is motivated by three assumptions Jansen (2001):

- The decorrelating property of a DWT creates a sparse signal, where most coefficients are zero or close to zero.
- Noise is spread out equally over all coefficients and the important signal singularities are still distinguishable from the noise coefficients.
- The noise level is not too high, so that we can recognize the signal wavelet coefficients.

2.2 Wavelet-based denoising

Basically, the problem of wavelet denoising by thresholding can be stated as follows. Let $\mathbf{g} = \left\{ g_{i,j}; i,j = 1,2,\ldots,M \right\}$ denotes the $M \times M$ observed image corrupted by additive Gaussian noise:

$$g_{i,j} = f_{i,j} + n_{i,j} \tag{24}$$

where $f_{i,j}$ is the noise-free pixel, $n_{i,j}$ has a $N(0,\sigma^2)$ distribution and σ^2 is the noise variance. Then, considering the linearity of the DWT:

$$y_{j,k} = x_{j,k} + z_{j,k} \tag{25}$$

with $y_{j,k}$, $x_{j,k}$ and $z_{j,k}$ denoting the k-th wavelet coefficient from the j-th decomposition level of the observed image, original image and noise image, respectively. The goal is to recover the unknown wavelet coefficients $x_{j,k}$ from the observed noisy coefficients $y_{j,k}$. One way to estimate $x_{j,k}$ is through Bayesian inference, by adopting a MAP approach. In this chapter, we introduce a MAP-MRF iterative method based on the combinatorial optimization algorithm *Game Strategy Approach* (GSA) Yu & Berthod (1995a), an alternative to the deterministic and widely known Besag's *Iterated Conditional Modes* (ICM) Besag (1986a). By iterative method we mean that an initial solution $\mathbf{x}^{(0)}$ is given and the algorithm successively improves it, by using the output from one iteration as the input to the next. Thus, the algorithm updates the current wavelet coefficients given a previous estimative according to the following MAP criterion:

$$\hat{x}_{j,k}^{(p+1)} = \boldsymbol{arg\,max}_{x_{j,k}} \left\{ p\left(x_{j,k} | x_{j,k}^{(p)}, y_{j,k}, \vec{\Psi} \right) \right\} \tag{26}$$

where $p\left(x_{j,k}|x_{j,k}^{(p)}, y_{j,k}, \vec{\Psi}\right)$ represents the *a posteriori* probability obtained by adopting a Generalized Gaussian distribution as likelihood (model for observations) and a Generalized Isotropic Multi-Level Logistic (GIMLL) MRF model as *a priori* knowledge (for contextual modeling), $x_{j,k}^{(p)}$ denotes the wavelet coefficient at p-th iteration and $\vec{\Psi}$ is the model parameter vector. This vector contains the parameters that control the behavior of the probability laws. More details on the statistical modeling and how these parameters are estimated are shown in Sections 3 and 4. In the following, we will derive an algorithm for approximating the MAP estimator by iteratively updating the wavelet coefficients.

3. The MAP-MRF framework for bayesian inference

The main problem with MAP-MRF approaches is that there is no analytical solution for MAP estimation. Hence, algorithms for numerically approximating the MAP estimator are required. It has been shown, in combinatorial optimization theory, that convergence to the global maximum of the posterior distribution can be achieved by the *Simulated Annealing* (SA) algorithm Geman & Geman (1984). However, as SA is extremely time consuming and demands a high computational burden, sub-optimal combinatorial optimization algorithms, which yield computationally feasible solutions to MAP estimation, are often used in real problems. Some of the most popular iterative algorithms found in image processing literature are: the widely used Besag's *Iterated Conditional Modes* (ICM) Besag (1986a), *Maximizer of the Posterior Marginals* (MPM) Marroquin et al. (1987a), *Graduated Non-Convexity* (GNC) Blake & Zisserman (1987), *Highest Con□dence First* (HCF) Chou & M. (1990) and *Deterministic Pseudo Annealing* Berthod et al. (1995). In this chapter, we introduce *GSAShrink*, a modified version of an alternative algorithm known as *Game Strategy Approach* (GSA) Yu & Berthod (1995a), based on non-cooperative game theory concepts and originally proposed for solving MRF image labeling problems.

3.1 Statistical modeling

3.1.1 Generalized gaussian distribution

It has been shown that the distribution of the wavelet coefficients within a sub-band can be modeled by a Generalized Gaussian (GG) with zero mean Mallat (1989), Westerink et al. (1991). The zero mean GG distribution has the probability density function:

$$p\left(w|\nu,\beta\right) = \frac{\nu}{2\beta\Gamma\left(1/\nu\right)} exp\left\{-\left(\frac{|w|}{\beta}\right)^{\nu}\right\} \tag{27}$$

where $\nu > 0$ controls the shape of the distribution and β the spread. Two special cases of the GG distribution are the Gaussian and the Laplace distributions. When $\nu = 2$ and $\beta = \sqrt{2}\sigma$, it becomes a standard Gaussian distribution. The Laplace distribution is obtained by setting $\nu = 1$ and $\beta = 1/\lambda$. According to Sharifi & Leon-Garcia (1995), the parameters ν and β can be empirically determined by directing computing the sample moments $\chi = E\left[|w|\right]$ and $\psi = E\left[w^2\right]$ (method of moments), because of this useful relationship:

$$\frac{\psi}{\chi^2} = \frac{\Gamma\left(\frac{1}{\tilde{\nu}}\right)\Gamma\left(\frac{3}{\tilde{\nu}}\right)}{\Gamma^2\left(\frac{2}{\tilde{\nu}}\right)} \tag{28}$$

and we can use a look-up table with different values of ν and determine is value from the ratio ψ/χ^2. After, it is possible to obtain $\hat{\beta}$ by:

$$\hat{\beta} = \frac{\psi\Gamma\left(\frac{1}{\tilde{\nu}}\right)}{\Gamma\left(\frac{3}{\tilde{\nu}}\right)} \tag{29}$$

3.1.2 Generalized isotropic multi-level logistic

Basically, MRF models represent how individual elements are influenced by the behavior of other individuals in their vicinity (neighborhood system). MRF models have proved to be powerful mathematical tools for contextual modeling in several image processing applications. In this chapter, we adopt a model originally proposed in Li (2009) that generalizes both Potts and standard isotropic Multi-Level Logistic (MLL) MRF models for continuous random variables. According to the Hammersley-Clifford theorem any MRF can be equivalently defined by a joint Gibbs distribution (global model) or by a set of local conditional density functions (LCDF's). From now on, we will refer to this model as Generalized Isotropic MLL MRF model (GIMLL). Due to our purposes and also for mathematical tractability, we define the following LCDF to characterize this model, assuming the wavelet coefficients are quantized into $\tilde{M}$ levels:

$$p\left(x_s|\eta_s,\theta\right) = \frac{\exp\left\{-\theta D_s\left(x_s\right)\right\}}{\sum_{y\in G}\exp\left\{-\theta D_s\left(y\right)\right\}} \tag{30}$$

where $D_s(y) = \sum_{k\in\eta_s}\left[1 - 2exp\left(-\left(y - x_k\right)^2\right)\right]$, x_s is the s-th element of the field, η_s is the neighborhood of x_s, x_k is an element belonging to the neighborhood of x_s, θ is a parameter that controls the spatial dependency between neighboring elements, and G is the set of all possible values of x_s, given by $G = \{g \in \Re / m \leq g \leq M\}$, where m and M are respectively, the minimum and maximum sub-band coefficients, with $|G| = \tilde{M}$ (cardinality of the set). This model provides a probability for a given coefficient depending on the similarity between its value and the neighboring coefficient values. Acording to Li (2009), the motivation for this model is that it is more meaningful in texture representation and easier to process than the isotropic MLL model, since it incorporates similarity in a softer way.

For GIMLL MRF model parameter estimation we adopt a Maximum Pseudo-Likelihood (MPL) framework that uses the observed Fisher information to approximate the asymptotic variance of this estimator, which provides a mathematically meaningful way to set this regularization parameter based on the observations. Besides, the MPL framework is useful in assessing the accuracy of MRF model parameter estimation.

3.2 Game strategy approach

In a n-person game, $I = \{1, 2, \ldots, n\}$ denotes the set of all players. Each player i has a set of pure strategies S_i. The game process consists in, at a given instant, each player choosing a strategy $s_i \in S_i$. Hence, a situation (or play) $\mathbf{s} = (s_1, s_2, \ldots, s_n)$ is yielded, and a payoff $H_i(\mathbf{s})$ is assigned to each player. In the approach proposed by Yu & Berthod (1995a), the payoff $H_i(\mathbf{s})$ of a player is defined in such a way that it depends only on its own strategy and on the set of strategies of neighboring players.

In non-cooperative game theory each player tries to maximize his payoff by choosing his own strategy independently. In other words, it is the problem of maximizing the global payoff through local and independent decisions, similar to what happens in MAP-MRF applications with the conditional independence assumption.

A mixed strategy for a player is a probability distribution defined over the set of pure strategies. In GSA, it is supposed that each player knows all possible strategies, as well as the payoff given by each one of them. Additionally, the solutions for a non-cooperative n-person game are given by the set of points satisfying the Nash Equilibrium condition (or *Nash points*). It has been shown that Nash Equilibrium points always exist in non-cooperative n-person games Nash (1950). A play $\mathbf{t}^* = \left(t_1^*, t_2^*, \ldots, t_n^*\right)$ satisfies the Nash Equilibrium condition if none of the players can improve you payoff by changing his strategy unilaterally, or in mathematical terms:

$$\forall i : H_i\left(\mathbf{t}^*\right) = max_{s_i \in S_i} H_i\left(\mathbf{t}^* || \mathbf{t}\right) \tag{31}$$

where $\mathbf{t}^* || \mathbf{t}$ is the play obtained by replacing $\mathbf{t}^*$ by $\mathbf{t}$.
The connection between game theory and combinatorial optimization algorithms is demonstrated in Yu & Berthod (1995a). It has been proved that the GSA algorithm fundamentals are based on two major results that states the equivalence between MAP-MRF estimation and non-cooperative games Yu & Berthod (1995a):

Theorem 3.1. (MAP-MRF Nash Points Equivalence Theorem) *The set of local maximum points of the a posteriori probability in MAP-MRF image labeling problems is identical to the set of Nash equilibrium points of the corresponding non-cooperative game.*

Theorem 3.2. (GSA Convergence Theorem) *The GSA algorithm converges to a Nash point in a finite number of iterations, given an arbitrary initial condition.*

Actually, a complete analogy between game theory and the wavelet denoising problem can be made, since the wavelet denoising process can be thought as being a generalization of image labeling, where instead of discrete labels, the unknown coefficients are continuous random variables. In Table 1 we show how concepts of non-cooperative game theory and our algorithm are in fact closely related.

Wavelet Denoising	Game Theory
sub-band lattice	n-person game structure
sub-band elements	players
wavelet coefficients	pure strategies
an entire sub-band at p-th iteration	a play or situation
posterior distribution	payoff
local conditional densities	mixed strategies
local maximum points (MAP)	Nash equilibrium points

Table 1. Correspondence between concepts of game theory and the MAP-MRF wavelet denoising approach.

3.3 GSAShrink for wavelet denoising

Given the observed data $\mathbf{y}$ (noisy image wavelet coefficients), and the estimated parameters for all the sub-bands $\vec{\Psi}_r = \left\{\hat{v}_r, \hat{\beta}_r, \hat{\theta}_r\right\}, r = 1, \ldots, S$, where S is the total number of sub-bands in the decomposition, our purpose is to recover the optimal wavelet coefficient field $\mathbf{x}^*$ using a Bayesian approach. As the number of possible candidates for $\mathbf{x}^*$ is huge, to make the problem computationally feasible, we adopt an iterative approach, where the wavelet coefficient field

at a previous iteration, let's say $\mathbf{x}^{(p)}$, is assumed to be known. Hence, the new wavelet coefficient $x_{j,k}^{(p+1)}$ can be obtained by:

$$x_{j,k}^{(p+1)} = \boldsymbol{argmax}_{x_{j,k}} \left\{ log\ p\left(x_{j,k}|\mathbf{x}^{(p)}, y_{j,k}, \vec{\Psi}_j\right)\right\} \tag{32}$$

Basically, *GSAShrink* consists in, given an initial solution, improve it iteratively by scanning all wavelet coefficients sequentially until the convergence of the algorithm or until a maximum number of iterations is reached. In this manuscript, we are setting the initial conditions as the own noisy image wavelet sub-band, that is, $\mathbf{x}^{(0)} = \mathbf{y}$, although some kind of previous preprocessing may provide better initializations. Considering the statistical modeling previously described, we can define the following approximation:

$$log\ p\left(x_{j,k}|\mathbf{x}^{(p)}, y_{j,k}, \vec{\Psi}_j\right) \propto log\left(\frac{\hat{v}_j}{2\hat{\beta}_j\Gamma\left(\frac{1}{\hat{v}_j}\right)}\right) - \tag{33}$$

$$\left[\frac{\left|y_{j,k}\right|}{\hat{\beta}_j}\right]^{\hat{v}_j} - \hat{\theta}_j \sum_{(\ell\in\eta_{j,k})}\left[1 - 2exp\left(-\left(x_{j,k}^{(p)} - x_{j,\ell}^{(p)}\right)^2\right)\right]$$

Therefore, we can define the following rule for updating the wavelet coefficient $x_{j,k}^{(p)}$, based on minimizing the negative of each player payoff, denoted by $H_{j,k}\left(\mathbf{x}, \mathbf{y}, \vec{\Psi}_j\right)$, considering $\mathbf{x}^{(0)} = \mathbf{y}$:

$$x_{j,k}^{(p+1)} = \boldsymbol{argmin}_{x_{j,k}} \left\{ H_{j,k}\left(\mathbf{x}, \mathbf{y}, \vec{\Psi}_j\right)\right\} \tag{34}$$

where

$$H_{j,k}\left(\mathbf{x}, \mathbf{y}, \vec{\Psi}_j\right) = \tag{35}$$

$$\left[\frac{\left|x_{j,k}\right|}{\hat{\beta}_j}\right]^{\hat{v}_j} + \hat{\theta}_j \sum_{(\ell\in\eta_{j,k})}\left[1 - 2exp\left(-\left(x_{j,k}^{(p)} - x_{j,\ell}^{(p)}\right)^2\right)\right]$$

The analysis of the above functional (the payoff of each player), reveals that while the first term favors low valued strategies (coefficients near zero), since the mean value of wavelet coefficients in a sub-band is zero, the MRF term favors strategies that are similar to those belonging to the neighborhood (coefficients close to the neighboring ones), defining a tradeoff between supression and smoothing, or hard and soft thresholding. In this scenario, the MRF model parameter plays the role of a regularization parameter, since it controls the compromisse between these two extreme behavior. Thus, our method can be considered a hybrid adaptive approach since identical wavelet coeficients belonging to different regions of a given sub-band are modified by completely different rules. In other words, coefficients belonging to smooth regions tend to be more attenuated than those belonging to coarser regions. In the following, we present the *GSAShrink* algorithm for wavelet-based image denoising.

ALGORITHM: GSAShrink for wavelet denoising

Require: The S sub-bands of the wavelet decomposition (LH_1, HL_1, HH_1, ...), a payoff function ($H_{j,k}$), the probability of acceptance of new strategies (α), the attenuation parameter for noisy coefficients (β), the gain parameter for relevant image coefficients (γ), the threshold (T) and the number of iterations (MAX).

Ensure: Shrinked wavelet sub-bands

while $p \leq MAX$ **do**
  for $j = 1$ to S **do**
    for $k = 1$ to $L(j)$ **do** {$L(j)$: size of current sub-band}
      $x_{j,k}^{*} = argmin_{x_{j,k}} \left\{ H_{j,k}\left(\mathbf{x}, \mathbf{y}, \vec{\Psi}_j\right) \right\}$
      if $\left(H\left(x_{j,k}^{*}\right) \leq H\left(x_{j,k}^{(p)}\right)\right)$ **then**
        if $\left(\left|x_{j,k}^{(p)}\right| \geq T\right)$ or $\left(max\left\{\left|\eta_{j,k}\right|\right\} \geq T\right)$ **then**
          $x_{j,k}^{(p+1)} = x_{j,k}^{(p)} \times (1+\gamma)$
        else
          $x_{j,k}^{(p+1)} = x_{j,k}^{*}$ w. p. α;
          Otherwise,
            $x_{j,k}^{(p+1)} = x_{j,k}^{(p)} \times (1-\beta)$ w. p. $(1-\alpha)$;
        end if
      end if
    end for
  end for
end while

It is interesting to note that an observation can be set forward to explain why there are a large number of "small" coefficients but relatively few "large" coefficients as the GGD suggests: the small ones correspond to smooth regions in a image and the large ones to edges, details or textures Chang et al. (2000). Therefore, the application of the derived MAP-MRF rule in all sub-bands of the wavelet decomposition removes noise in an adaptive manner by smoothing the wavelet coeficients in a selective way.

Basically, the *GSAShrink* algorithm works as follows: for each wavelet coefficient, the value that maximizes the payoff is chosen and the new payoff is calculated. If this new payoff is less than the original one, then nothing is done (since in the Nash equilibrium none of the playes can improve its payoff by uniterally changing its strategy). Otherwise, if the absolute value of the current wavelet coefficient $x_{j,k}$ or any of its neighbors is above the threshold T, which means that we are probably dealing with relevant image information such as edges or fine details, then $x_{j,k}$ is amplified by a factor of $(1+\gamma)$. The goal of this procedure is to perform some image enhancement during noise removal. However, if its magnitude is less a threshold, then the new coefficient $x_{j,k}^{*}$ is accepted with probability α, which is a way to smooth the wavelet coefficients since we are employing the MAP-MRF functional given by equation (35). The level of suppression/shrinkage depends basically on two main issues: the contextual information and the MRF model parameter, that controls the tradeoff between suppression and smoothing. On the other hand, with probability $(1-\alpha)$ the coefficient is attenuated by a constant factor of $(1-\beta)$, since we are probably facing a noise coefficient. It is worthwhile to note that the only parameter originally existing in the traditional GSA algorithm for image labeling problems is α, that controls the probability of acceptance of new strategies. Both β

and γ parameters have been included to better represent the nature of our problem. Also, in all experiments thoughout this chapter, we have adopted the following parameter values: $\alpha = 0.9$, $\beta = 0.1$, $\gamma = 0.05$ and $MAX = 5$.

3.4 Wavelet thresholds

As we have seen, a critical issue in the method is the choice of the thresholding value. Several works in the wavelet literature discuss threshold estimation Chang et al. (2000); Jansen & Bultheel (1999). In the experiments throughout this chapter we adopted four different wavelet thresholdings: universal Donoho (1995); Donoho et al. (1995), SURE Jansen (2001), Bayes and Oracle thresholds Chang et al. (2000).

3.4.1 Universal threshold

Despite its simplicity, it has been shown that the Universal Threshold has some optimal asymptotic properties Donoho (1995); Donoho & Johnstone (1994). The Universal Threshold is obtained by the following expression:

$$\lambda_{UNIV} = \sqrt{2logN\sigma^2} \tag{36}$$

where N is the number of data points and σ^2 denotes the noise variance. Thus, the Universal Threshold does not depend directly on the observed input signal, but only on simple statistics derived from it.

3.4.2 SURE threshold

The SURE (*Steins's Unbiased Risk Estimator*) threshold is obtained by minimizing a risk function $R(.)$, assuming the coefficients are normally distributed Hudson (1978); Stein (1981). In this chapter, we use the approximation for $R(.)$ derived in Jansen (2001) and given by:

$$R(\lambda) = \left[\frac{1}{N}\left\|\omega_\lambda - \omega\right\|^2 - \sigma^2\right] + \left[2\sigma^2\frac{(N - N_0)}{N}\right] \tag{37}$$

where N is the number of wavelet coefficients, σ^2 is the noise variance, ω_λ and ω denote the wavelet coefficients before and after thresholding, respectively, and N_0 is number of null wavelet coefficients after thresholding. The SURE threshold λ_{SURE}, is defined as the one that minimizes $R(\lambda)$, that is:

$$\lambda_{SURE} = \boldsymbol{argmin}_\lambda \left\{R(\lambda)\right\} \tag{38}$$

Analyzing the expression we can see that this method for threshold estimation seeks a tradeoff between data fidelity and noise removal.

3.4.3 Bayes threshold

The Bayes Threshold is a data-driven threshold derived in Bayesian framework by using a generalized Gaussian distribution as prior for the wavelet coefficients. It is a simple and closed-form threshold that is obtained in a sub-band adaptive way by Chang et al. (2000):

$$\lambda_{BAYES} = \frac{\hat{\sigma}^2}{\hat{\sigma}_x} \tag{39}$$

where

$$\hat{\sigma}_x = \sqrt{max\left(\hat{\sigma}_y^2 - \hat{\sigma}^2, 0\right)} \tag{40}$$

$$\hat{\sigma}_y^2 = \frac{1}{N^2}\sum_{1}^{N} y_i^2 \tag{41}$$

$$\hat{\sigma} = \frac{Median\left(|y_i|\right)}{0.6745} \tag{42}$$

It is worth mentioning that in case of $\hat{\sigma}^2 > \hat{\sigma}_y^2$, $\hat{\sigma}_x$ is taken to be zero, implying that $\lambda_{BAYES} = \infty$, which means, in practice, that all coefficients within the sub-band are suppressed.

3.4.4 Oracle thresholds

The Oracle Thresholds are the theoretic optimal sub-band adaptive thresholds in a MSE sense, assuming the original image is known, a condition that obviously is possible only in simulations. The *OracleShrink* threshold is defined as:

$$\lambda_S^* = \boldsymbol{argmin}_\lambda \left\{ \sum_{k=1}^{N} \left(\eta_\lambda\left(y_k\right) - x_k\right)^2 \right\} \tag{43}$$

where N is the number of wavelet coefficients in the sub-band, η_λ denotes the soft thresholding operator and x_k is the *k-th* coefficient of the original image. Similarly, the *OracleThresh* threshold is given by:

$$\lambda_H^* = \boldsymbol{argmin}_\lambda \left\{ \sum_{k=1}^{N} \left(\psi_\lambda\left(y_k\right) - x_k\right)^2 \right\} \tag{44}$$

where ψ_λ denotes the hard threshold operator.

4. Statistical inference on MRF models

With advances on probability and statistics, such as the remarkable Hammersley-Clifford Theorem Hammersley & Clifford (1971), which states the Gibbs-Markov equivalence, Bayesian inference and the development of Markov Chain Monte Carlo simulation methods (MCMC) Metropolis et al. (1953), Geman & Geman (1984), Swendsen & Wang (1987), Wolff (1989) together with combinatorial optimization algorithms to solve numerical maximization of complex high dimensional functions Besag (1986b), Marroquin et al. (1987b), Yu & Berthod (1995b), Markov Random Fields became a central topic in a variety of research fields including pattern recognition, game theory, computer vision and image processing. Those important contributions have led to a huge number of novel methodologies and techniques in statistical applications, especially those regarding contextual modeling and spatial data analysis.
However, in most applications the MRF model parameters are still chosen by trial-and-error through simple manual adjustments Solberg (2004), Wu & Chung (2007). Therefore, statistical inference on several MRF models remains an open problem. The main reason is that the most general estimation method, maximum likelihood (ML), is computationally intractable. An alternative solution proposed by Besag (1974) is to use the local conditional density functions (LCDF) to perform Maximum Pseudo-Likelihood (MPL) estimation.

4.1 Maximum pseudo-likelihood estimation

This section briefly describes the MLP estimation of the Generalized isotropic MLL parameter model θ, given by equation (30). Basically, our motivations for this approach are:

- MPL estimation is a computationally feasible method.
- From a statistical perspective, MPL estimators have a series of desirable properties, such as consistency and asymptotic normality Jensen & Künsh (1994), Winkler (2006).

In recent works found in MRF literature, analytical pseudo-likelihood equations for Potts MRF model on higher-order neighborhood systems have been derived Levada et al. (2008c), showing the importance of MRF parameter estimation assessment. In the experiments along this chapter, the proposed methodology is based on the approximation for the asymptotic variance of the Potts MRF model reported in Levada et al. (2008b) and Levada et al. (2008a).

4.1.1 Pseudo-likelihood equation

The main advantage of maximum pseudo-likelihood estimation is its mathematical tractability and computational simplicity. The pseudo-likelihood function for the Generalized Potts MRF model is defined as:

$$PL(X;\theta) = \prod_{s=1}^{N} \frac{\exp\{-\theta D_s(x_s)\}}{\sum_{y\in G}\exp\{-\theta D_s(y)\}} \tag{45}$$

where N denotes the number of elements on the field.
Taking the logarithms, differentiating on the parameter and setting the result to zero, lead to the following expression (pseudo-likelihood equation):

$$\sum_{s=1}^{N}\left[\frac{\sum_{y\in G} D_s(y)\, exp\{-\theta D_s(y)\}}{\sum_{y\in G} exp\{-\theta D_s(y)\}}\right] = \sum_{s=1}^{N} D_s(x_s) \tag{46}$$

In the experiments, the solution is obtained by finding the zero of the resultant equation. We chose the Brent's method Brent (1973), a numerical algorithm that does not require the computation (or even the existence) of derivatives. The advantages of this method are: it uses a combination of bisection, secant, and inverse quadratic interpolation methods, leading to a very robust approach. Besides, it has superlinear convergence rate.

4.2 Bilateral filtering

Bilateral Filtering (BF) is a noniterative and local non-linear spatial domain filtering technique that originally was proposed as an intuitive tool Tomasi & Manduchi (1998) but later has showed to be closely related to classical partial differential equation based methods, more precisely, anisotropic diffusion Barash (2002); Dong & Acton (2007); Elad (2002). The basic idea of bilateral filtering is to use a weighted average of degraded pixels to recover the original pixel by combining a low-pass function (h_D) and a edge stoping function (h_P) according to the following relationship:

$$\hat{f}[i,j] = \frac{\sum_{(k,n)\in\Omega_{i,j}} h_D[k,n]h_P[k,n]g(k,n)}{\sum_{(k,n)\in\Omega_{i,j}} h_D[k,n]h_P[k,n]} \tag{47}$$

where $\Omega_{i,j}$ is a $(2N+1)\times(2N+1)$ window centered at (i,j) and

$$h_D[k,n] = exp\left\{-\frac{(k-i)^2+(n-j)^2}{2\sigma_D^2}\right\} \tag{48}$$

$$h_P[k,n] = exp\left\{-\frac{(g[k,n]-g[i,j])^2}{2\sigma_P^2}\right\} \tag{49}$$

where the parameters σ_D and σ_P control the effect of the spatial and radiometric weight factors. The first weight, h_D, measures the geometric distance between the central pixel and each one of its neighbors, in a way that the nearest samples have more influence on the final result than the distant ones. The second weight, h_P, penalizes the neighboring pixels that vary greatly in intensity from the central pixel, in a way that the larger the difference, the smaller will be the pixel's contribution during the smoothing. In all experiments along this chapter, we set $N = 2$ (5×5 window), $\sigma_D^2 = 1$ and $\sigma_P^2 = 0.1$.

Basis	**Metrics**			
		Soft	Hard	*GSAShrink*
HAAR	ISNR	-0.8484	0.4388	0.5823
	PSNR	25.613	27.032	27.777
	SSIM	0.8012	0.8625	**0.8903**
DB4	ISNR	0.4864	1.6952	2.2662
	PSNR	27.067	28.705	29.365
	SSIM	0.8598	0.9017	**0.9108**
SYM4	ISNR	0.6580	1.8093	2.3455
	PSNR	27.257	28.662	29.266
	SSIM	0.8639	0.9001	**0.9113**
BIOR6.8	ISNR	0.8336	1.9868	2.587
	PSNR	27.549	28.856	29.829
	SSIM	0.8655	0.8981	**0.9176**

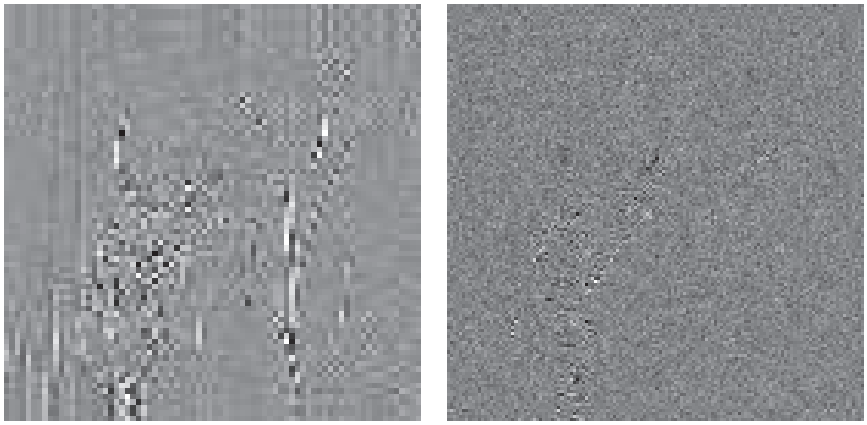
Table 2. Performance of wavelet denoising algorithms on Lena image corrupted by additive Gaussian noise (PSNR = 26.949 dB) using the Universal Threshold.

5. Experiments and results

In order to test and evaluate the *GSAShrink* algorithm for wavelet-based image denoising, we show the results of some experiments performed by using both simulated and real noisy image data:

- Lena image corrupted by gaussian noise.
- Real Nuclear Magnetic Resonance (NMR) images from primate brains (marmosets and brown capuchin monkeys).

In all experiments the wavelet thresholds were estimated in a sub-band adaptive way, which means that we used a different threshold λ_j, $j = 1, 2, \ldots, 6$, for each sub-band, except the LL_2 (approximation), since we are using a Level-2 wavelet decomposition, resulting in the six details sub-bands known as LL_2, LH_2, HL_2,HH_2, LH_1,HL_1 and HH_1. Also, in all experiments, we compared the performance of *GSAShrink* against *soft* and *hard-thresholding* techniques, by using several wavelet basis: *Haar*, *Daubechies4*, *Symlet4* and *Biorthogonal6.8*, a kind of



(a) LHL wavelet subband. (b) HH wavelet subband.

Fig. 2. HL_2 and HH_1 wavelet sub-bands for the Lena image: (a) a more homogeneous situation ($\theta = 1.1754$) and (b) a more heterogeneous case ($\theta = 0.9397$), defined by statistically different MRF parameter values.

wavelet transform that has filters with symmetrical impulse response, that is, linear phase filters. The motivation for including Biorthogonal wavelets is that it has been reported that in image processing applications filters with non-linear phase aften introduce visually annoying artifacts in the denoised images.

To perform quantitative analysis of the obtained results, we compared several metrics for image quality assessment. In this manuscript, we selected three different metrics that are: Improvement in Signal-To-Noise-Ratio (ISNR), Peak Signal-To-Noise Ratio (PSNR) and Structural Similarity Index (SSIM), since MSE based metrics have proved to be inconsistent with the human eye perception Wang & Bovik (2009).

Sub-band	$\hat{\theta}_{MPL}$	$\hat{\sigma}_n^2(\hat{\theta}_{MPL})$
LH_2	1.1441	3.1884×10^{-6}
HL_2	1.1754	9.1622×10^{-6}
HH_2	1.0533	1.8808×10^{-5}
LH_1	0.9822	6.2161×10^{-6}
HL_1	0.9991	7.3409×10^{-6}
HH_1	0.9397	4.5530×10^{-6}

Table 3. MPL estimators for θ and asymptotic variances for the Lena image wavelet sub-bands.

Table 2 shows the results for *GSAShrink* denoising on the Lena image, corrupted by additive Gaussian noise ($PSNR = 26.949$ dB). Table 3 shows the estimated regularization MRF parameters and their respective asymptotic variances for each one of the details sub-bands. Figure 2 shows the HL_2 and HH_1 sub-bands of wavelet decomposition. Note that the coarser a sub-band, the smaller is the regularization parameter, indicating that suppression is favored over smoothing, forcing a more intense noise removal.

Analyzing the results, we see that *GSAShrink* had superior performance in all cases. Furthermore, the best result was obtained by using *GSAShrink* with Biorthogonal 6.8 wavelets. To illustrate these numerical results, Figure 3 shows some visual results for the best performances.

(a) Noisy Lena. (b) Soft-Threshold.

(c) Hard Threshold. (d) GSAShrink.

Fig. 3. Visual results for wavelet denoising using Biorthogonal6.8 wavelets with sub-band adaptive Universal threshold (Table 2): (a) Noisy Lena; (b) Soft-Threshold; (c) Hard-Threshold; (d) GSAShrink.

The same experiment was repeated by considering other threshold estimation methods. The use of SURE and Bayes thresholds improved the denoising performance, as indicate Table 4. As the use of Biothogonal6.8 wavelets resulted in uniformly superior performances, from now on we are omitting the other wavelet filters. Figure 4 shows the visual results for the best results (SURE).

As *GSAShrink* iterativelly converges to local maxima solutions, we performed an experiment to illustrate the effect of using different initializations on the final result by combining spatial domain (*Bilateral Filtering*) and wavelet-domain (*GSAShrink*) non-linear filtering. If instead of considering the observed noisy image directly as input to our algorithm, we use the result of *Bilateral Filtering*, the performance can be further improved. Table 5 shows a comparison between simple *Bilateral Filtering* and the combined approach. Figure 5 shows that the use of

Threshold	Metrics			
		Soft	Hard	*GSAShrink*
SURE	ISNR	2.0235	2.8836	3.3458
	PSNR	28.702	29.641	30.441
	SSIM	0.8918	0.8991	0.9270
Bayes	ISNR	2.8511	1.2721	3.2280
	PSNR	29.433	28.270	29.880
	SSIM	0.8942	0.8306	0.9157
Oracle	ISNR	3.3713	2.7318	3.6411
	PSNR	30.045	29.586	30.609
	SSIM	0.9103	0.8964	0.9277

Table 4. Performance of wavelet denoising algorithms on Lena image corrupted by additive Gaussian noise (PSNR = 26.949 dB) using the SURE, Bayes and Oracle thresholds with Biorthogonal6.8 wavelets.

(a) Noisy Lena. (b) Soft-Threshold.

(c) Hard-Threshold. (d) GSAShrink.

Fig. 4. Visual results for wavelet denoising using Biorthogonal6.8 wavelets with sub-band adaptive SURE threshold (Table 4): (a) Noisy Lena; (b) Soft-Threshold; (c) Hard-Threshold; (d) GSAShrink.

Bilateral Filtering in the generation of initial conditions to the *GSAShrink* algorithm prevents the appearance of visible artifacts that are usually found in wavelet-based methods.

Metric	BF	*GSAShrink*	BF + *GSAShrink*
ISNR	4.6211	3.2280	4.6912
PSNR	31.149	29.880	31.481
SSIM	0.9142	0.9157	0.9310

Table 5. Results of using Bilateral Filtering to generate better initial conditions to our MAP-MRF approach.

Sub-band	$\hat{\theta}_{MPL}$	$\hat{\sigma}_n^2(\hat{\theta}_{MPL})$
LH_2	0.8066	2.4572×10^{-5}
HL_2	0.8898	3.7826×10^{-5}
HH_2	0.7338	1.0153×10^{-5}
LH_1	0.7245	3.9822×10^{-5}
HL_1	0.7674	5.8675×10^{-5}
HH_1	0.6195	4.4578×10^{-5}

Table 6. MPL estimators for θ and asymptotic variances for the NMR image wavelet sub-bands.

5.1 Results on real image data

Additionally to the simulations, we have performed some experiments on real NMR image data to test and evaluate *GSAShrink*. The NMR images considered here are from primate brains (both marmosets and brown capuchin monkeys) and were acquired by the CInAPCe project, an abbreviation for the Portuguese expression "Inter-Institutional Cooperation to Support Brain Research" a Brazilian research project that has as main purpose the establishment of a scientific network seeking the development of neuroscience research through multidisciplinary approaches. In this sense, image processing can contribute to the development of new methods and tools for analyzing magnetic resonance imaging and its integration with other methodologies in the investigation of brain diseases.

Figure 6 shows some visual results for NMR image denoising. Analyzing the results, it is possible to see that *GSAShrink* acts more like *soft-thresholding* in homogeneous areas and more like *hard-thresholding* in regions with a lot of details. Table 6 shows the estimated regularization MRF parameters and their respective asymptotic variances for each one of the details sub-bands. Figure 7 shows the subbands HL_2 and HH_1.

6. Conclusion

In this chapter, we investigated a novel MAP-MRF iterative algorithm for wavelet-based image denoising (*GSAShrink*). Basically, it uses the Bayesian approach and game-theoretic concepts to build a flexible and general framework for wavelet shrinkage. Despite its simplicity, *GSAShrink* has demonstrated to be efficient in edge-preserving image filtering. The Generalized Gaussian distribution and a GIMLL MRF model were combined to derive a payoff function which provides a rule for iterativelly update the current value of a wavelet coefficient. This was, to the best of our knowledge, the first time these two models were combined for this purpose. Also, we have shown that in this scenario, the MRF model parameter plays the same role of a regularization parameter, since it controls the tradeoff between supression and atenuation, defining a hybrid approach.

Experiments with both simulated and real NMR image data provided good results that were validated by several quantitative image quality assessment metrics. The obtained results

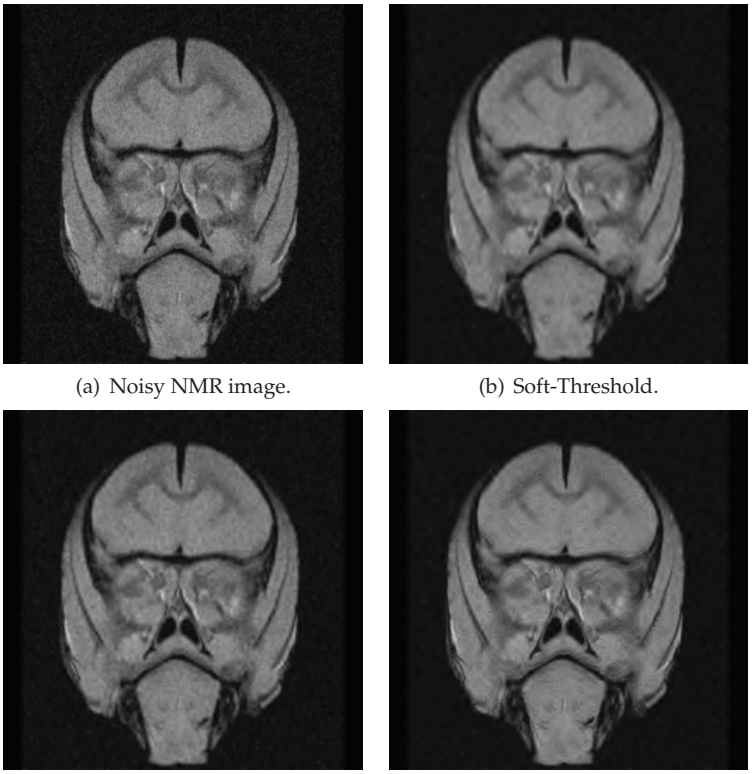
(a) Original Lena. (b) Bilateral Filtering (BF).

(c) GSAShrink. (d) BL + GSAShrink.

Fig. 5. Results for wavelet denoising using combination of *Bilateral Filtering* and our MAP-MRF approach (Table 5): (a) Original Lena; (b) Bilateral Filtering (BF); (c) GSAShrink; (d) Bilateral Filtering + GSAShrink.

indicated a significant improvement in the denoising performance, showing the efectiveness of the proposed method.

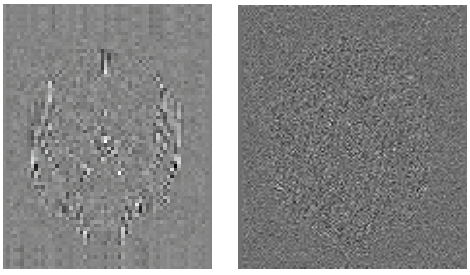
Future works may include the use and investigation of more wavelet decomposition levels, other kinds of wavelet transforms, such as wavelet packets and undecimated or stationary transforms, as well as the filtering of other kinds of noise such as multiplicative speckle and signal-dependent Poisson noise (by using the Anscombe Transform). Finally, we intend to proposed and study the viability of other combinatorial optimization shrinkage methods as *ICMShrink* and *MPMShrink*, based on modified versions of ICM and MPM algorithms respectively. Regarding the influence of the initial conditions on the final result, we believe that the use of multiple initializations instead of a single one, together with information fusion techniques, can further improve the denoising performance, particularly in multiframe image filtering/restoration or video denoising, where several frames from the same scene are available and only the noise changes from one frame to another.



(a) Noisy NMR image. (b) Soft-Threshold.

(c) Hard-Threshold. (d) BL + GSAShrink.

Fig. 6. Results for wavelet denoising on real NMR marmoset brain image data: (a) Noisy NMR image; (b) Soft-Threshold; (c) Hard-Threshold; (d) Bilateral Filtering + GSAShrink.



(a) LHL subband. (b) HH subband.

Fig. 7. HL_2 and HH_1 wavelet sub-bands for the NMR image: (a) a more homogeneous situation ($\theta = 0.8898$) and (b) a more heterogeneous case ($\theta = 0.6195$), defined by statistically different MRF parameter values.

7. References

Barash, D. (2002). A fundamental relationship between bilateral filtering, adaptive smoothing and the nonlinear diffusion equation, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24(6): 844–847.

Berthod, M., Kato, Z. & Zerubia, J. (1995). Dpa: Deterministic approach to the map problem, *IEEE Transactions on Image Processing* 4(9): 1312–1314.

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems, *Journal of the Royal Statistical Society - Series B* 36: 192–236.

Besag, J. (1986a). On the statistical analysis of dirty pictures, *Journal of the Royal Statistical Society B* 48(3): 192–236.

Besag, J. (1986b). On the statistical analysis of dirty pictures, *Journal of Royal Statistical Society Series B* 48(3): 259–302.

Blake, A. & Zisserman, A. (1987). *Visual Reconstruction*, MIT Press.

Brent, R. (1973). *Algorithms for minimization without derivatives*, Prentice Hall.

Chang, S. G., Yu, B. & Vetterli, M. (2000). Adaptive wavelet thresholding for image denoising and compression, *IEEE Trans. on Image Processing* 9(9): 1532–1546.

Chou, P. B. & M., B. C. (1990). The theory and practice of bayesian image labeling, *International Journal of Computer Vision* 4: 185–210.

Daubechies, I. (1988). Orthonormal bases of compactly supported wavelets, *Communications on Pure and Applied Mathematics* 41(7): 909–996.

Dong, G. & Acton, S. T. (2007). On the convergence of bilateral filter for edge-preserving image smoothing, *IEEE Signal Processing Letters* 14(9): 617–620.

Donoho, D. L. (1995). De-noising by soft-thresholding, *IEEE Trans. on Information Theory* 41(3): 613–627.

Donoho, D. L. & Johnstone, I. M. (1994). Ideal spatial adaptation via wavelet shrinkage, *Biometrika* 81: 425–455.

Donoho, J., Johnstone, I. M., Kerkyacharian, G. & Picard, D. (1995). Wavelet shrinkage: Asymptopia ?, *Journal of the Royal Statistical Society B* 52(2): 301–369.

Elad, M. (2002). On the origin of the bilateral filtering and ways to improve it, *IEEE Transactions on Image Processing* 11(10): 1141–1151.

Geman, S. & Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images, *IEEE Trans. on Pattern Analysis Machine Intelligence* 6(6): 721–741.

H., Y., Zhao, L. & Wang, H. (2009). Image denoising using trivariate shrinkage filter in the wavelet domain and joint bilateral filter in the spatial domain, *IEEE Transactions on Image Processing* 18(10): 2364–2369.

Hammersley, J. & Clifford, P. (1971). Markov field on finite graphs and lattices. unpublished.

Hudson, H. M. (1978). A natural identity for exponential families with applications in multiparameter estimation, *Annals of Statistics* 6(3): 473–484.

Jansen, A. & Bultheel, A. (1999). Multiple wavelet threshold estimation by generalized crossvalidation for images with correlated noise, *IEEE Transactions on Image Processing* 8(7): 947–953.

Jansen, M. (2001). *Noise reduction by wavelet thresholding*, Springer-Verlag.

Jensen, A. & Cour-Harbo, A. (2001). *Ripples in Mathematics*, Springer-Verlag Berlin.

Jensen, J. & Künsh, H. (1994). On asymptotic normality of pseudo likelihood estimates for pairwiseinteraction processes, *Annals of the Institute of Statistical Mathematics* 46(3): 475–486.

Ji, H. & Fermüller, C. (2009). Robust wavelet-based super-resolution reconstruction: Theory and algorithm, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 31(4): 649–660.

Levada, A. L. M., Mascarenhas, N. D. A. & Tannús, A. (2008a). A novel pseudo-likelihood equation for potts mrf model parameter estimation in image analysis, INTERNATIONAL CONFERENCE ON IMAGE PROCESSING, ICIP, 15., 2008, San Diego. *Proceedings...*, IEEE, San Diego/CA, pp. 1828–1831.

Levada, A. L. M., Mascarenhas, N. D. A. & Tannús, A. (2008b). On the asymptotic variances of gaussian markov random field model hyperparameters in stochastic image modeling, INTERNATIONAL CONFERENCE ON PATTERN RECOGNITION, ICPR, 19., 2008, Tampa. *Proceedings...*, IEEE, Tampa/FL, pp. 1–4.

Levada, A., Mascarenhas, N. & Tannús, A. (2008c). Pseudolikelihood equations for potts mrf model parameter estimationon higher-order neighborhood systems, *IEEE Geoscience and Remote Sensing Letters* 5(3): 522–526.

LI, S. Z. (2009). *Markov Random Field Modeling in Image Analysis*, 3 edn, Springer-Verlag London, Inc.

Mallat, S. G. (1989). A theory of multiresolution image decomposition: The wavelet representation, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 11(7): 647–693.

Marroquin, J., Mitter, S. & Poggio, T. (1987a). Probabilistic solution of ill-posed problems in computer vision, *Journal of American Statistical Society* 82: 76–89.

Marroquin, J., Mitter, S. & Poggio, T. (1987b). Probabilistic solution of ill-posed problems in computer vision, *Journal of American Statistical Society* 82: 76–89.

Metropolis, N., Rosenbluth, A., Rosenbluth, M. & Teller, A.and Teller, E. (1953). Equation of state calculations by fast computer machines, *Journal of Physical Chemistry* 21: 1987–2092.

Nash, J. F. (1950). Equilibrium points in n-person games, *Proceedings of the National Academy of Sciences* 36: 48–49.

Nasri, M. & Nezamabadi-pour, H. (2009). Image denoising in the wavelet domain using a new adaptive thresholding function, *Neurocomputing* 72: 1012–1025.

Sharifi, K. & Leon-Garcia, A. (1995). Estimation of shape parameter for generalized gaussian distributions in sub-band decompositions of video, *IEEE Transactions on Circuits and Systems for Video Technology* 5(1): 52–56.

Solberg, A. H. S. (2004). Flexible nonlinear contextual classification, *Pattern Recognition Letters* 25: 1501–1508.

Stein, C. (1981). Estimation of the mean of a multivariate normal distribution, *Annals of Statistics* 9(6): 1135–1151.

Strang, G. & Nguyen, T. (1997). *Wavelets and Filter Banks*, Wellesley-Cambridge Press.

Swendsen, R. & Wang, J. (1987). Nonuniversal critical dynamics in monte carlo simulations, *Physical Review Letters* 58: 86–88.

Tomasi, C. & Manduchi, R. (1998). Bilateral filtering for gray and color images, INTERNATIONAL CONFERENCE ON COMPUTER VISION, ICCV, 6., 1998, Bombay. *Proceedings...*, IEEE, Bombay, India, pp. 839–846.

Wang, Z. & Bovik, A. C. (2009). Mean squared error: Love it or leave it ? a new look at signal fidelity measures, *IEEE Signal Processing Magazine* 26(1): 98–117.

Westerink, P. H., Biemond, J. & Boekee, D. E. (1991). *Sub-band Image Coding*, Kluwer Academic, chapter Sub-band coding of color images.

Winkler, G. (2006). *Image Analysis, Random Fields and Markov Chain Monte Carlo Methods: A Mathematical Introduction*, Springer.

Wolff, U. (1989). Collective monte carlo updating for spin systems, *Physical Review Letters* 62: 361–364.

Wu, J. & Chung, A. C. S. (2007). A segmentation model using compound markov random fields based ona boundary model, *IEEE Transactions on Image Processing* 16(1): 241–252.

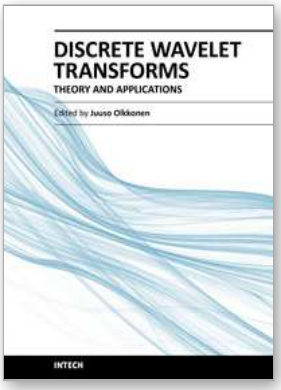
Yoon, B. J. & Vaidyanathan, P. P. (2004). Wavelet-based denoising by customized thresholding, *Proceedings of International Conference on Acoustics, Speech, and Signal Processing*, Vol. 2, pp. 925–928.

Yu, S. & Berthod, M. (1995a). A game strategy approach for image labeling, *Computer Vision and Image Understanding* 61(1): 32–35.

Yu, S. & Berthod, M. (1995b). A game strategy approach for image labeling, *Computer Vision and Image Understanding* 61(1): 32–37.

Zhang, B. & Allebach, J. P. (2008). Adaptive bilateral filter for sharpness enhancement and noise removal, *IEEE Transactions on Image Processing* 17(5): 664–678.

Zhang, M. & Gunturk, B. K. (2008). Multiresolution bilateral filtering for image denoising, *IEEE Transactions on Image Processing* 17(12): 2324–2333.



Discrete Wavelet Transforms - Theory and Applications

Edited by Dr. Juuso T. Olkkonen

ISBN 978-953-307-185-5
Hard cover, 256 pages
Publisher InTech
Published online 04, April, 2011
Published in print edition April, 2011

Discrete wavelet transform (DWT) algorithms have become standard tools for discrete-time signal and image processing in several areas in research and industry. As DWT provides both frequency and location information of the analyzed signal, it is constantly used to solve and treat more and more advanced problems. The present book: Discrete Wavelet Transforms: Theory and Applications describes the latest progress in DWT analysis in non-stationary signal processing, multi-scale image enhancement as well as in biomedical and industrial applications. Each book chapter is a separate entity providing examples both the theory and applications. The book comprises of tutorial and advanced material. It is intended to be a reference text for graduate students and researchers to obtain in-depth knowledge in specific applications.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Alexandre L. M. Levada, Nelson D. A. Mascarenhas and Alberto Tannús (2011). A MAP-MRF Approach for Wavelet-Based Image Denoising, Discrete Wavelet Transforms - Theory and Applications, Dr. Juuso T. Olkkonen (Ed.), ISBN: 978-953-307-185-5, InTech, Available from: http://www.intechopen.com/books/discrete-wavelet-transforms-theory-and-applications/a-map-mrf-approach-for-wavelet-based-image-denoising

INTECH
open science | open minds

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821