GSAShrink: A Novel Iterative Approach for Wavelet-Based Image Denoising

Alexandre L. M. Levada[1], Alberto Tannús[3]
Instituto de Física de São Carlos
Universidade de São Paulo
São Carlos/SP, Brasil
alexandreluis@ursa.ifsc.usp.br, goiano@ifsc.usp.br

Nelson D. A. Mascarenhas[2]
Departamento de Computação
Universidade Federal de São Carlos
São Carlos/SP, Brasil
nelson@dc.ufscar.br

Abstract—In this paper we propose a novel iterative algorithm for wavelet-based image denoising following a *Maximum a Posteriori* (MAP) approach. The wavelet shrinkage problem is modeled according to the Bayesian paradigm, providing a strong and extremely flexible framework for solving general image denoising problems. To approximate the MAP estimator, we propose *GSAShrink*, a modified version of a known combinatorial optimization algorithm based on non-cooperative game theory (*Game Strategy Approach*, or GSA). In order to modify the original algorithm to our purposes, we generalize GSA by introducing some additional control parameters and steps to reflect the nature of wavelet shrinkage applications. To test and evaluate the proposed method, experiments using several wavelet basis on noisy images are proposed. Additionally to better visual quality, the obtained results produce quantitative metrics (MSE, PSNR, ISNR and UIQ) that show significant improvements in comparison to traditional wavelet denoising approaches known as soft and hard thresholding, indicating the effectiveness of the proposed algorithm.

Keywords-Image Denoising; Wavelets; Bayesian Estimation; Maximum a Posteriori; Game Strategy Approach

I. INTRODUCTION

Image denoising is a required pre-processing step in most image processing, computer vision and pattern recognition applications, since in many situations they may be corrupted by noise during acquisition and transmission stages. Therefore, estimating a signal that is degraded by additive noise has been of interest to a wide community of researchers. Basically, the goal of denoising is to remove the noise as much as possible, while retaining important image features, such as edges and finer details. Traditional denoising methods are based on linear filtering, where the most usual choices are Wiener and convolutional filters. Lately, a vast literature on non-linear filtering has emerged, especially those based on wavelets [1],[2], [3] inspired by the remarkable work of Donoho [4].

In the wavelet-based noise reduction method called wavelet shrinkage, the wavelet coefficients of a noisy image are divided into important and non-important coefficients and each one of these groups are modified by certain rules. Usually, in most denoising applications *soft* and *hard* thresholding are considered. Basically, filtering is performed by comparing each wavelet coefficient to a given threshold and setting it to zero (hard) or attenuating it (soft) if

its magnitude is less than the threshold; otherwise, it is kept untouched. *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 [5], [4]. 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 [1] conclude that the optimal *soft-thresholding* estimator yields a smaller estimation error than the optimal *hard-thresholding* estimator.

However, for some classes of signals, *hard-thresholding* results in superior estimates to that of *soft-thresholding*, despite some of its disadvantages [2]. To tackle this problem, several different thresholding functions are proposed in the literature. The basic idea consists in defining custom thresholding functions, which are similar to hard thresholding, but having a smooth transition around the threshold [3]. Examples of this kind of custom thresholding functions are *semisoft* or *firm shrinkage* and *non-negative garrote* thresholding function [6]. In [2], a custom function that can be viewed as a linear combination of *hard-thresholding* function and *soft-thresholding* function is proposed. Results of the application of this kind of functions usually show better performance when compared with traditional *hard* and *soft-thresholding* techniques.

In this work, we propose *GSAShrink*, a novel adaptive algorithm for wavelet-based image denoising following a Bayesian approach. Basically, there are three main motivations for the proposed method. First, to achieve results that are not so smooth as the ones provided by *soft-thresholding* neither so coarse as the results of *hard-thresholding*. We want to smooth homogeneous regions, keeping relevant image information as edges and finer details. In this sense, the Bayesian inference is an excellent solution, since it provides a robust and extremely flexible mathematical framework for stochastic image modeling. The fundamental principle of Bayesian philosophy is to regard all observable variables as unknown stochastic quantities, assigning probability distributions based on our subjective beliefs and prior knowledge

[7]. Second, traditional wavelet thresholding methods operates on one single coefficient at a time, completely ignoring its neighborhood. Contextual information can be useful in determining how a given coefficient should be modified. Our choice is to incorporate a Markov Random Field (MRF) model, since MRF's are powerful mathematical tools for contextual modeling in image processing applications [8], [9], [10]. Finally, the use of an iterative approach has several advantages over traditional one-step methods, as the convergence control, the incorporation of local information (or heuristics) according to the nature of the problem and also, in this specific case, the fact that a given wavelet coefficient can be updated in more than one way all along the iterative process, depending on its current status (adaptive method).

To test and evaluate the proposed method, we built a series of experiments on simulated image data (Lena image), using several wavelet bases. The obtained results show the effectiveness of *GSAShrink*, indicating a clear improvement on the wavelet denoising performance in comparison to the traditional approaches.

The remaining of the paper is organized as follows. Section 2 briefly introduces the discrete wavelet transform (DWT) and the wavelet denoising problem. In Section 3 we discuss the proposed method, as well as the statistical modeling and the estimation of the necessary parameters. Section 4 is concerned about metrics for image quality assessment and performance evaluation. The experiments and results are described in Section 5. Finally, Section 6 brings the conclusions and final remarks.

II. THE WAVELET TRANSFORM

The wavelet transform is a mathematical tool that decomposes a given signal in a basis of orthogonal functions. However, differently from the Fourier Transform, which uses periodic, smooth and unlimited basis functions (sines and cosines), DWT uses *wavelets*, that is, non-periodic, non-smooth and finite support basis functions, allowing a much more meaningful representation through multi-resolution analysis. 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) $h[]$, a low-pass filter, and $g[]$, the corresponding high-pass filter.

A. Wavelet Denoising by Thresholding

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

$$g_{i,j} = f_{i,j} + n_{i,j} \tag{1}$$

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:

$$y_{j,k} = x_{j,k} + z_{j,k} \tag{2}$$

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. We propose a novel iterative method based on the combinatorial optimization algorithm GSA. 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 proposed method updates the current wavelet coefficients given a previous estimative according to the following MAP criterion:

$$\hat{x}_{j,k}^{(p+1)} = \boldsymbol{arg\ max}_{x_{j,k}} \left\{ p\left(x_{j,k}|x_{j,k}^{(p)}, y_{j,k}, \vec{\Psi} \right) \right\} \tag{3}$$

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 the observations) and a MRF model called Generalized Isotropic Multi-Level Logistic (GIMLL) 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 Section (3.2). In the following, we will derive an algorithm for approximating the MAP estimator by iteratively updating the wavelet coefficients.

III. ALGORITHMS FOR BAYESIAN INFERENCE

Now that the wavelet denoising is stated as a Bayesian inference problem, algorithms for 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 *Simulated Annealing* (SA). However, as SA is extremely time consuming and demands a high computational cost, sub-optimal combinatorial optimization algorithms, which yield computationally feasible solutions to MAP estimation, were proposed. Some of the most popular combinatorial optimization algorithms found in image processing literature are: the widely recognized *Iterated Conditional Modes* (ICM) [11], *Maximizer of the Posterior Marginals* (MPM) [12], *Graduated Non-Convexity* (GNC) [13], *Highest Confidence First* (HCF) [14] and *Deterministic Pseudo Annealing* [15]. In this work, we propose a modified version of an alternative algorithm, named *Game Strategy Approach* (GSA) [16],

based on non-cooperative game theory, originally proposed for solving MRF image labeling problems.

A. 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 [16], 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 defined as 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 [17]. A play $\mathbf{t}^* = (t_1^*, t_2^*, \ldots, t_n^*)$ satisfies the Nash Equilibrium condition if none of the players can improve you payoff by changing his strategy unilaterally, or in mathematical terms:

$$\forall i : H_i(\mathbf{t}^*) = max_{s_i \in S_i} H_i(\mathbf{t}^*||\mathbf{t}) \qquad (4)$$

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 [16]. Actually, still according to [16], the GSA fundamentals are based on two main propositions:

PROPOSITION 1. *The set of local maximum points of the a posteriori probability in MRF image labeling problems is identical to the set of Nash equilibrium points of the corresponding non-cooperative game.*

PROPOSITION 2. *The GSA relaxation algorithm converges to a Nash equilibrium when the number of iterations increases.*

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

Table I
CORRESPONDENCE BETWEEN CONCEPTS OF GAME THEORY AND THE PROPOSED WAVELET DENOISING APPROACH.

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

B. Statistical Modeling

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 [18], [19]. The zero mean GG distribution has the probability density function:

$$p(w|\nu, \beta) = \frac{\nu}{2\beta\Gamma(1/\nu)} exp\left\{-\left(\frac{|w|}{\beta}\right)^\nu\right\} \qquad (5)$$

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 [20], the parameters ν and β can be empirically determined by directing computing the sample moments $\chi = E[|w|]$ and $\psi = E[w^2]$ (method of moments), because of this useful relationship:

$$\frac{\psi}{\chi^2} = \frac{\Gamma\left(\frac{1}{\nu}\right)\Gamma\left(\frac{3}{\nu}\right)}{\Gamma^2\left(\frac{2}{\nu}\right)} \qquad (6)$$

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:

$$\hat{\beta} = \frac{\psi\Gamma\left(\frac{1}{\nu}\right)}{\Gamma\left(\frac{3}{\nu}\right)} \qquad (7)$$

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). In this work, we adopt a model originally proposed in [9] that generalizes the standard isotropic Multi-Level Logistic (MLL) MRF model for continuous random fields. According to the Hammersley-Clifford theorem a MRF can be defined by a Gibbs joint distribution or by a set of local conditional density functions (LCDF's). We will call this model Generalized Isotropic MLL MRF model (GIMLL). Due to our purposes and also for mathematical

tractability, we define the following LCDF to characterize this model.

$$p\left(x_s|\eta_s,\theta\right) = \frac{\exp\left\{-\theta D_s\left(x_s\right)\right\}}{\sum_{y\in G}\exp\left\{-\theta D_s\left(y\right)\right\}} \quad (8)$$

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 spatial dependency parameter that controls how the central element is influenced by its neighbors, and G is the set of all possible values of x_s, given by $G = \{g/m \leq g \leq M\}$, where m and M are respectively the minimum and maximum sub-band coefficients. This model provides a probability for a given coefficient depending on the similarity between its value and the neighboring coefficient values.

For the estimation of the θ parameter in each sub-band, we adopt Maximum Pseudo-Likelihood (MPL) estimation. The main advantage of MPL estimation is its mathematical tractability and computational simplicity. The pseudo-likelihood function for the GIMLL model is defined as:

$$PL(X;\theta) = \prod_{s=1}^{N} p\left(x_s|\eta_s,\theta\right) =$$
$$= \prod_{s=1}^{N}\frac{\exp\left\{-\theta D_s(x_s)\right\}}{\sum_{y\in G}\exp\left\{-\beta D_s(y)\right\}} \quad (9)$$

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

$$\frac{\partial}{\partial\theta}logPL(X;\theta) = -\sum_{s=1}^{N} D_s\left(x_s\right) + \quad (10)$$

$$+\sum_{s=1}^{N}\left[\frac{\sum_{y\in G} D_s\left(y\right)exp\left\{-\hat{\theta}D_s\left(y\right)\right\}}{\sum_{y\in G}exp\left\{-\hat{\theta}D_s\left(y\right)\right\}}\right] = 0$$

In the experiments, the solution is obtained by finding the zero of the resultant equation. We chose Brent's method, a numerical algorithm that does not require the computation (or even the existence) of derivatives. The advantages of this algorithm 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.

C. GSAShrink for wavelet denoising

Given the observed data $\mathbf{y}$ (noisy image wavelet coefficients), and the estimated parameters for all the sub-bands $\vec{\Psi}_r = \left\{\hat{\nu}_r,\hat{\beta}_r,\hat{\theta}_r\right\}$, $r = 1,\ldots,S$, 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:

$$x_{j,k}^{(p+1)} = \boldsymbol{argmax}_{x_{j,k}}\left\{log\ p\left(x_{j,k}|\mathbf{x}^{(p)},y_{j,k},\vec{\Psi}_j\right)\right\} \quad (11)$$

Basically, the proposed method 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 work, 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:

$$log\ p\left(x_{j,k}|\mathbf{x}^{(p)},y_{j,k},\vec{\Psi}_j\right) \propto log\left(\frac{\hat{\nu}_j}{2\hat{\beta}_j\Gamma\left(\frac{1}{\hat{\nu}_j}\right)}\right) - \quad (12)$$

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

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 $\mathbf{x}^{(0)} = \mathbf{y}$:

$$x_{j,k}^{(p+1)} = \boldsymbol{argmin}_{x_{j,k}}\left\{H_{j,k}\left(\mathbf{x},\mathbf{y},\vec{\Psi}_j\right)\right\} \quad (13)$$

where

$$H_{j,k}\left(\mathbf{x},\mathbf{y},\vec{\Psi}_j\right) = \quad (14)$$

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

The analysis of the above functional (the payoff of each player), reveals that while the first term favors the appearance of 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 the nearest neighbor players (coefficients close to the neighboring ones), defining a kind of regularization procedure. 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 [1]. Therefore, application of the proposed method in all sub-bands of the wavelet decomposition leads to smoother version of the image, since it attenuates the noise. In the following, we present the algorithm for *GSAShrink* for wavelet-based image denoising.

Algorithm 1: GSASHRINK FOR WAVELET SHRINKAGE

INPUT: *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 (N).*

OUTPUT: *Shrinked wavelet sub-bands.*

1.   Repeat for every wavelet sub-band, while $p \leq N$.
2.   For each wavelet coefficient of the sub-band
3.   Choose the coefficient $x_{j,k}^*$ that minimizes the negative of the payoff:
4. $x_{j,k}^* = argmin_{x_{j,k}} \left\{ H_{j,k} \left(\mathbf{x}, \mathbf{y}, \vec{\Psi}_j \right) \right\}$
5.   If $\left(H \left(x_{j,k}^* \right) \leq H \left(x_{j,k}^{(p-1)} \right) \right)$
6.   If $\left(\left| x_{j,k}^* \right| \geq T \right)$ or $\left(max \left\{ \left| \eta_{j,k} \right| \right\} \geq T \right)$
7. $x_{j,k}^{(p)} = x_{j,k}^{(p-1)} \times (1 + \gamma)$
8.   Else
9.   Accept $x_{j,k}^*$ w. p. α;
10.   Otherwise,
11. $x_{j,k}^{(p)} = x_{j,k}^{(p-1)} \times (1 - \beta)$ w. p. $(1 - \alpha)$;
12.   END
13.   END
14.   END

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 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 $(1 + \gamma)$. However, if $|x_{j,k}| \leq T$, then we accept the new coefficient with probability α, or attenuate the coefficient by a factor of $(1 - \beta)$ with probability $(1 - \alpha)$, since its likely that we are dealing with noise information (low valued coefficient). The only parameter existing in the original GSA algorithm for image labeling is α. In this work, we adopt $\alpha = 0.9$, $\beta = 0.01$, $\gamma = 0.1$ and $N = 10$.

D. Metrics for Image Quality Assessment

In order to perform quantitative analysis of the obtained results, we compare several metrics for image quality assessment. In this work, we selected four different metrics that are: the traditional Mean Square Error (MSE), Improvement in Signal-To-Noise-Ratio (ISNR) and Peak Signal-To-Noise Ratio (PSNR) and Universal Image Quality Index (UIQ), a metric that takes perceptual fidelity into account using a combination of three factors: loss of correlation, luminance distortion, and contrast distortion [21], given by:

$$UIQ \left(\mathbf{x}, \mathbf{y} \right) = \frac{4\mu_x \mu_y \sigma_{xy}}{\left(\mu_x^2 + \mu_y^2 \right) \left(\sigma_x^2 + \sigma_y^2 \right)} \qquad (15)$$

where μ_x and μ_y are the mean values of original and filtered images, σ_x and σ_y are the standard deviations of original and filtered images, and σ_{xy} is the sample cross-correlation.

For an excellent discussion on how challenging is to measure image quality and also the drawbacks and advantages of each measure, as well as a complete definition of each metric, the reader is referred to [22].

IV. EXPERIMENTS AND RESULTS

In order to test and evaluate the proposed *GSAShrink* method for wavelet-based image denoising, we performed some experiments using noisy image data. In the first experiment, we compared the performance of the proposed method against *soft* and *hard-thresholding* techniques, by using several wavelet basis: *Haar*, *Daubechies4*, *Symlet4* and *Biorthogonal6.8*, a kind of wavelet transform that has filters with symmetrical impulse response, that is, linear phase filters. The motivation for including this basis is that it has been reported that in image processing applications filters with non-linear phase can introduce artifacts that are visually annoying.

In all experiments we considered the adaptive Universal Threshold, which means that we calculated a different threshold T_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 sub-bands known as LL_2, LH_2, HL_2, HH_2, LH_1, HL_1 and HH_1. The Universal Threshold is calculated by $T_{univ} = \sqrt{2logN}\sigma$, where N is the number of coefficients of the sub-band and σ^2 is the noise variance. Table II shows the results for *GSAShrink* denoising on the Lena image, corrupted by additive Gaussian noise $MSE = 131.268$, $PSNR = 26.949$ dB.

Note that in all situations the proposed method outperforms both *soft* and *hard-thresholding* techniques. Figures 1 and 2 shows the noisy image and the results of wavelet denoising using *Biorthogonal6.8* wavelets. As the quantitative metrics suggests, the *GSAShrink* denoised image presents a significant better visual quality. Note also that *GSAShrink* provides a result that is smoother and with less artifacts than *hard-thresholding* does, but at the same time preserving much more high frequency content such as edges and details in comparison to *soft-thresholding*.

Despite the improvement in the image quality, the proposed method has a limitation regarding the computational cost. In our experiments, *GSAShrink* performed about 50 times slower than traditional one-step methods. Such difference in computational time is due to the iterative nature of the proposed method, since it has to be applied to all wavelet sub-bands in the decomposition, sequentially.

Also, the maximization procedure demands the calculation of the payoff of each player for all possible strategies, increasing the computational complexity. While traditional methods take just a few seconds (average execution time: 10 seconds), *GSAShrink* can take minutes to perform wavelet-based image denoising (average execution time: 500 seconds). However, an alternative to attenuate this problem would be the use of parallel programming techniques, since coefficients within a wavelet sub-band do not depend on coefficients from different sub-bands.

(a) Noisy Lena

(b) Soft Thresholding

(c) Hard Thresholding

(d) *GSAShrink*

Figure 1. Visual results for wavelet denoising using Biorthogonal6.8 wavelets with sub-band adaptive Universal threshold (Table II).

(a) Noisy Lena

(b) Soft Thresholding

(c) Hard Thresholding

(d) *GSAShrink*
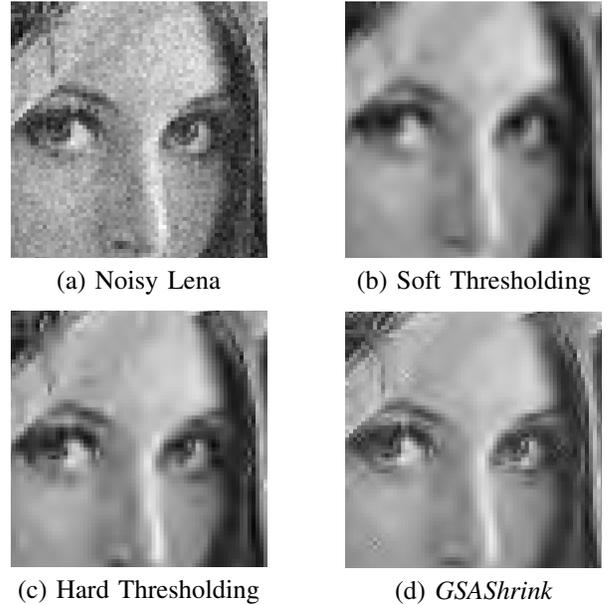
Figure 2. Visual results for wavelet denoising using Biorthogonal6.8 wavelets with sub-band adaptive Universal threshold (Table II) (Zoom in).

Table II
PERFORMANCE OF WAVELET DENOISING ALGORITHMS ON LENA IMAGE CORRUPTED BY ADDITIVE GAUSSIAN NOISE (PSNR = 26.949 DB).

Basis	Metrics	Soft	Hard	*GSAShrink*
HAAR	MSE	159.947	118.512	114.707
	ISNR	-0.8484	0.4388	0.5823
	PSNR	25.613	27.032	27.777
	UIQ	0.9365	0.9546	0.9609
		Soft	Hard	*GSAShrink*
DB4	MSE	117.268	88.778	77.157
	ISNR	0.4864	1.6952	2.2662
	PSNR	27.067	28.705	29.365
	UIQ	0.9509	0.9647	0.9717
		Soft	Hard	*GSAShrink*
SYM4	MSE	112.725	86.475	75.551
	ISNR	0.6580	1.8093	2.3455
	PSNR	27.257	28.662	29.266
	UIQ	0.9533	0.9656	0.9723
		Soft	Hard	*GSAShrink*
BIOR6.8	MSE	108.258	83.012	71.325
	ISNR	0.8336	1.9868	2.587
	PSNR	27.549	28.856	29.829
	UIQ	0.9543	0.9666	0.9731

In order to demonstrate the quality of the *GSAShrink* results, we compared our results with the truly optimal sub-band adaptive thresholds in a MSE sense for *soft* and *hard-thresholding*, assuming the original image is known: *OracleShrink* and *OracleThresh*, proposed by [1]. The *OracleShrink* threshold is defined as:

$$T_{OS}^* = \textbf{\textit{argmin}}_T \left\{ \sum_{k=1}^{N} \left(\eta_T (y_k) - x_k \right)^2 \right\} \quad (16)$$

where N is the number of wavelet coefficients in the sub-band and η_T denotes the soft thresholding operator. Similarly, the *OracleThresh* threshold is given by:

$$T_{OT}^* = \textbf{\textit{argmin}}_T \left\{ \sum_{k=1}^{N} \left(\psi_T (y_k) - x_k \right)^2 \right\} \quad (17)$$

where ψ_T denotes the hard threshold operator. Table III shows the metrics for *OracleShrink*, *OracleThresh* and *GSAShrink* on several wavelet basis. Note that all MSE-based metrics show the superiority of *GSAShrink* over both soft and hard Thresholding. Figure 3 shows a comparison

between the visual results for *OracleShrink* and *GSAShrink* using the theoretic optimum minimum MSE threshold. The obtained results indicate that the proposed method is capable of achieving better performances in all situations, by filtering the noise, but at the same time preserving fine and relevant image details.

(a) *B6.8 + OracleShrink*

(b) *B6.8 + GSAShrink*

Figure 3. Visual results for wavelet denoising using Biorthogonal6.8 with OracleShrink and GSAShrink + theoretic optimum minimum MSE Threshold (from Table III).

V. CONCLUSION

In this paper, we proposed a novel iterative algorithm for wavelet-based image denoising, named *GSAShrink*. Basically, it uses the Bayesian framework and Game Theory concepts to build a flexible and general approach for wavelet shrinkage. The Generalized Gaussian distribution and a MRF model are combined to derive a payoff function which gives a rule for updating the current value of a wavelet coefficient. Experiments with simulated data provided good results that were validated by several quantitative image quality assessment metrics. The obtained results indicated a significant improvement in the denoising performance, showing the efectiveness of the proposed method. Future works may include the use of other types of thresholds (*SUREShrink*, *BayesShrink*, etc.), more wavelet decomposition levels and

Table III
PERFORMANCE OF OPTIMAL MSE SOFT AND HARD THRESHOLDS ON LENA IMAGE CORRUPTED BY ADDITIVE GAUSSIAN NOISE (PSNR = 26.949 DB).

Basis	Metrics			
		OrShrink	*OrThresh*	*GSAShrink*
HAAR	MSE	75.389	103.587	76.665
	ISNR	2.4051	1.0252	2.3322
	PSNR	28.930	28.062	29.526
	UIQ	0.9716	0.9645	0.9735
		OrShrink	*OrThresh*	*GSAShrink*
DB4	MSE	64.630	76.329	59.008
	ISNR	3.0738	2.3513	3.4691
	PSNR	29.761	29.242	30.547
	UIQ	0.9767	0.9722	0.9784
		OrShrink	*OrThresh*	*GSAShrink*
SYM4	MSE	65.535	86.589	62.399
	ISNR	2.9921	1.8036	3.2060
	PSNR	29.687	28.738	30.007
	UIQ	0.9763	0.9695	0.9772
		OrShrink	*OrThresh*	*GSAShrink*
BIOR6.8	MSE	59.950	69.927	56.438
	ISNR	3.3713	2.7318	3.6462
	PSNR	30.045	29.586	30.729
	UIQ	0.9778	0.9736	0.9790

the filtering of other kinds of noise such as multiplicative speckle (by applying a logarithmic transformation to the image) and signal-dependent Poisson noise (by using the Anscombe Transform) in real images. Different statistical models are also an object of study, since each model leads to a specific payoff function. Also, the analysis of other image quality assessment measures is definitely a very important issue, since it has been shown that MSE often fails in predicting human perception of image fidelity and quality [22]. 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, as well as the use of wavelet packets.

REFERENCES

[1] S. G. Chang, B. Yu, and M. Vetterli, "Adaptive wavelet thresholding for image denoising and compression," *IEEE Trans. on Image Processing*, vol. 9, no. 9, pp. 1532–1546, September 2000.

[2] B. J. Yoon and P. P. Vaidyanathan, "Wavelet-based denoising by customized thresholding," in *Proceedings of International Conference on Acoustics, Speech, and Signal Processing*, vol. 2, 2004, pp. 925–928.

[3] M. Nasri and H. Nezamabadi-pour, "Image denoising in the wavelet domain using a new adaptive thresholding function," *Neurocomputing*, vol. 72, pp. 1012–1025, 2009.

[4] D. L. Donoho, "De-noising by soft-thresholding," *IEEE Trans. on Information Theory*, vol. 41, no. 3, pp. 613–627, 1995.

[5] D. L. Donoho and I. M. Johnstone, "Ideal spatial adaptation via wavelet shrinkage," *Biometrika*, vol. 81, pp. 425–455, 1994.

[6] H. Y. Gao and A. G. Bruce, "Wavelet shrinkage denoising using the non-negative garrote," *Journal of Computational and Graphical Statistics*, vol. 7, no. 4, pp. 469–488, 1998.

[7] A. K. Katsaggelos, R. Molina, and J. Mateos, *Super Resolution of Images and Video*. Morgan & Claypool, 2007.

[8] C. S. Won and R. M. Gray, *Stochastic Image Processing*. Springer-Verlag, 2004.

[9] S. Z. Li, *Markov Random Field Modeling In Image Analysis*. Springer-Verlag, 2001.

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

[11] J. Besag, "On the statistical analysis of dirty pictures," *Journal of the Royal Statistical Society B*, vol. 48, no. 3, pp. 192–236, 1986.

[12] J. Marroquin, S. Mitter, and T. Poggio, "Probabilistic solution of ill-posed problems in computer vision," *Journal of American Statistical Society*, vol. 82, pp. 76–89, 1987.

[13] A. Blake and A. Zisserman, *Visual Reconstruction*. MIT Press, 1987.

[14] P. B. Chou and B. C. M., "The theory and practice of bayesian image labeling," *International Journal of Computer Vision*, vol. 4, pp. 185–210, 1990.

[15] M. Berthod, Z. Kato, and J. Zerubia, "Dpa: Deterministic approach to the map problem," *IEEE Transactions on Image Processing*, vol. 4, no. 9, pp. 1312–1314, 1995.

[16] S. Yu and M. Berthod, "A game strategy approach for image labeling," *Computer Vision and Image Understanding*, vol. 61, no. 1, pp. 32–35, 1995.

[17] J. F. Nash, "Equilibrium points in n-person games," *Proceedings of the National Academy of Sciences*, vol. 36, pp. 48–49, 1950.

[18] S. G. Mallat, "A theory of multiresolution image decomposition: The wavelet representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 7, pp. 647–693, 1989.

[19] P. H. Westerink, J. Biemond, and D. E. Boekee, *Sub-band Image Coding*. Kluwer Academic, 1991, ch. Sub-band coding of color images.

[20] K. Sharifi and A. Leon-Garcia, "Estimation of shape parameter for generalized gaussian distributions in sub-band decompositions of video," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 5, no. 1, pp. 52–56, 1995.

[21] Z. Wang and A. C. Bovik, "A universal image quality index," *IEEE Signal Processing Letters*, vol. 9, no. 3, pp. 81–84, 2002.

[22] ——, "Mean squared error: Love it or leave it ? a new look at signal fidelity measures," *IEEE Signal Processing Magazine*, vol. 26, no. 1, pp. 98–117, 2009.