ADAPTIVE IMAGE DENOISING IN SCALE-SPACE USING THE WAVELET TRANSFORM

JUNG C.R.[1], SCHARCANSKI J.[2]

[1,2] UFRGS–Universidade Federal do Rio Grande do Sul
Instituto de Informática
Av. Bento Gonçalves, 9500. Porto Alegre, RS, Brasil 91501-970
[1] jung@inf.ufrgs.br
[2] jacobs@inf.ufrgs.br

[1] UNISINOS - Universidade do Vale do Rio dos Sinos
Centro de Ciências Exatas e Tecnológicas
Av. UNISINOS, 950. São Leopoldo, RS, Brasil, 93022-000
crjung@exatas.unisinos.br

Abstract. This paper proposes a new method for image denoising with edge preservation, based on image multiresolution decomposition by a redundant wavelet transform. In our approach, edges are implicitly located and preserved in the wavelet domain, while noise is filtered out. At each resolution, the coefficients associated to noise and coefficients associated to edges are modeled by Gaussians, and a shrinkage function is assembled. The shrinkage functions are combined in consecutive resolution, and geometric constraints are applied to preserve edges that are not isolated. Finally, the inverse wavelet transform is applied to the modified coefficients. This method is adaptive, and performs well for images contaminated by natural and artificial noise.

Keywords: denoising, edge detection, multiresolution analysis, wavelet shrinkage.

1 Introduction

An important and difficult problem in image processing is to remove noise from images without blurring the edges. Typically, noise is characterized by high spatial frequencies, and Fourier-based techniques tend to suppress high frequencies, which also smears the edges. On the other hand, the wavelet transform provides good localization in both spatial and spectral domains, allowing noise removal and edge preservation. Nowadays, there are several approaches based on the wavelet transform, with promising results.

The method proposed by Mallat and Hwang [1] estimates local regularity of the image by calculating the Lipschitz exponents. Coefficients with low Lipschitz exponent values are removed, and the image is reconstructed using the remaining coefficients (in fact, only the local maxima are used, and the image is estimated from these maxima). The reconstruction process is based on an iterative projection procedure, which may be computationally demanding (although Carmona developed a noninterative algorithm [2] that was shown to be equivalent to the iterative procedure, but much faster). This approach does not produce good results for images with low signal-to-noise ratios (SNR), and requires user interaction.

Xu *et al.* [3] used the correlation of wavelet coefficients between consecutive scales to distinguish noise from meaningful data. Their method is based on the fact that wavelet coefficients related to noise are less correlated across scales than coefficients associated to edges. If the correlation is smaller than a threshold, a given coefficient is set to zero. To determine a proper threshold, a noise power estimate is needed by their technique, which may be difficult to obtain for some images.

Malfait and Roose [4] developed a technique that takes into account two measures for image filtering. The first is a measure of local regularity of the image based on the Hölder exponent, and the second takes into account geometric constraints. These two measures are combined in a Bayesian probabilistic formulation, and implemented by a Markov random field model. The SNR gain achieved by this method is significant, but the stochastic sampling procedure needed for the probabilities calculation is computationally demanding.

Simoncelli and Adelson [5] proposed a probabilistic approach for image denoising using wavelets. The authors used a two-parameter Generalized Laplacian distribution to model the wavelet coefficients of the image. These parameters are estimated from noisy observations. The result is a nonlinear estimator that performs wavelet coefficient shrinkage.

Chang *et al.* [6, 7] proposed a probabilistic approach

1530-1834/01 $10.00 © 2001 IEEE

for image denoising using soft thresholds. In [6], the coefficients associated to noise are modeled as Gaussian functions, while coefficients associated to edges are modeled as Generalized Gaussian functions. These probability functions are used to determine a soft threshold. In a modification of their technique [7], context modeling was also utilized. Each coefficient is modeled as a random variable of a Generalized Gaussian distribution, based on spatial information (context modeling within a neighborhood), and a threshold is estimated for each coefficient.

Mihçak *et al.* [8] proposed a spatially adaptive statistical model for image denoising. The wavelet coefficients are modeled as Gaussian random variables with high local correlation, and a maximum *a posteriori* probability rule is applied to restore the original coefficients from the noisy observations.

Also, Strela *et al.* [9] described the joint densities of clusters of wavelet coefficients as a Gaussian scale mixture, and developed a maximum likelihood solution for estimating relevant wavelet coefficients from the noisy observations.

The main problem of the probabilistic approaches in [5, 6, 7, 8, 9] is that a noise estimate is needed, which may be difficult to obtain in practical applications, specially in images with inherent noise (such as aerial images, MRI images, etc.).

Pizurica *et al.* [10] proposed a computationally efficient method for image filtering, that utilizes local noise measurements and geometrical constraints in the wavelet domain. A *shrinkage function* based on these two measures is used to selectively modify the wavelet coefficients, and the image is reconstructed based on the updated wavelet coefficients. Although this method is fast, it does not take into account the evolution of wavelet coefficients along scales, which usually carries important information. Images with low signal-to-noise ratio (SNR) are not efficiently denoised by this technique.

This paper proposes a new method for image denoising using the wavelet transform, which combines wavelet shrinkage and consistency in scale-space. The wavelet transform with two detail images (horizontal and vertical) is computed, the distributions of these coefficients are modeled by a composition of Gaussian probability density functions, and a shrinkage function is assembled at each scale. The shrinkage functions for consecutive levels are then combined to preserve edges that are persistent in scale-space (i.e., appear in several consecutive scales), and geometric constraints are applied to remove residual noise.

The next section gives a brief description of the wavelet framework, and the section that follows describes the new method. Section 4 presents some experimental results obtained with our approach, and a comparison with other denoising techniques. Conclusions are presented in the final section.

2 The Wavelet Transform

In this paper, we use the wavelet decomposition with only two detail images (horizontal and vertical) [11], instead of the classical approach where three detail images are used (horizontal, vertical and diagonal details) [12]. Also, we use a non-subsampled filter bank with the mother wavelet proposed in [11]. This approach requires one smoothing function $\phi(x,y)$ and two wavelets $\psi^1(x,y)$, $\psi^2(x,y)$. The dilation of these functions are denoted by:

$$\phi_s(x,y) = \frac{1}{s^2}\phi(\frac{x}{s},\frac{y}{s}) \text{ and } \psi^i_s(x,y) = \frac{1}{s^2}\psi^i(\frac{x}{s},\frac{y}{s}), \tag{1}$$

and the dyadic wavelet transform $f(x,y)$, at a scale $s = 2^j$, has two detail components given by:

$$W^i_{2^j} f(x,y) = (f * \psi^i_{2^j})(x,y),\ i = 1,2, \tag{2}$$

and one low-pass component, given by:

$$S_{2^j} f(x,y) = (f * \phi_{2^j})(x,y). \tag{3}$$

The coefficients $W^1_{2^j} f(x,y)$ and $W^2_{2^j} f(x,y)$ represent the details in the x and y directions, respectively. Thus, the image gradient at the resolution 2^j can be approximated by:

$$\mathbf{W}_{2^j} f(x,y) = \begin{pmatrix} W^1_{2^j} f(x,y) \\ W^2_{2^j} f(x,y) \end{pmatrix}. \tag{4}$$

Since we are dealing with digital images $f[n,m]$, we use the discrete version of the wavelet transform [11], and denote the discrete wavelet coefficients by $W^i_{2^j} f[n,m]$, for $i = 1,2$.

The edge magnitudes at the resolution 2^j can be computed from:

$$M_{2^j} f[n,m] = \sqrt{(W^1_{2^j} f[n,m])^2 + (W^2_{2^j} f[n,m])^2}, \tag{5}$$

and the edge orientation is given by the gradient direction, which is expressed by:

$$\theta_{2^j} f[n,m] = \arctan\left(\frac{W^2_{2^j} f[n,m]}{W^1_{2^j} f[n,m]}\right). \tag{6}$$

3 Our Image Denoising Approach

The main idea of our technique is to modify the detail coefficients $W^1_{2^j} f[n,m]$ and $W^2_{2^j} f[n,m]$, so that coefficients associated to edges are kept, and coefficients associated to noise are removed. The inverse wavelet transform is applied to the modified wavelet coefficients, resulting in the denoised image.

Next we describe a technique that assigns to each coefficient a probability of being an edge, and propagates this information along the scale-space, using consistency along scales and geometric continuity.

3.1 Wavelet Shrinkage

Wavelet shrinkage is a known approach for noise reduction, where the wavelet coefficients are subject to a non-linearity that reduces (or suppresses) low-amplitude values and retains high-amplitude values [13, 5].

For each level 2^j, we want to find non-negative non-decreasing shrinkage functions $g_j^i(x)$, with $0 \leq g_j^i(x) \leq 1$, such that the wavelet coefficients $W_{2^j}^i f$ are updated according the following rule:

$$NW_{2^j}^i f[n,m] = W_{2^j}^i f[n,m] g_j^i(W_{2^j}^i f[n,m]), \quad i = 1,2, \tag{7}$$

where $g_j^i(W_{2^j}^i f[n,m])$ is a *shrinkage factor*. Let us denote $g_j^i[n,m] = g_j^i(W_{2^j}^i f[n,m])$. Since the same shrinkage procedure (Equation (7)) will be applied to all scales and subbands of the wavelet decomposition, the indexes j and i will be used only when necessary to remove ambiguity.

To estimate the shrinkage factor $g[n,m]$, we analyze the distribution of coefficients $Wf[n,m]$. Some of these coefficients are related to noise, and others to edges. If the image is contaminated by additive white noise, the corresponding coefficients $Wf[n,m]$ may be considered Gaussian distributed [14], with standard deviation σ_{noise}. Then, the Gaussian distribution function that models these coefficients is given by:

$$p(x|\text{noise}) = \frac{1}{\sigma_{\text{noise}}\sqrt{2\pi}} e^{-x^2/2\sigma_{\text{noise}}^2} \tag{8}$$

In practice, we observe that noise-free images typically consist of homogeneous regions and not many edges. In general, homogeneous regions contribute to a sharp peak near zero for the histograms of $Wf[n,m]$, and the edges contribute to the tail of the distribution. This distribution presents a sharper peak than a Gaussian [5], and therefore the Gaussian model is not appropriate for the distribution of the noise-free coefficients. However, we assume that the distribution of the wavelet coefficients $Wf[n,m]$ related exclusively to edges (and not related to homogeneous regions) can be approximated by a Gaussian (i.e., when the sharp peak in $Wf[n,m]$ associated to homogeneous regions is not considered, we assume that the remaining data is approximated by a Gaussian function). Therefore, the distribution of edge-related coefficients $Wf[n,m]$ is approximated by the function:

$$p(x|\text{edge}) = \frac{1}{\sigma_{\text{edge}}\sqrt{2\pi}} e^{-x^2/2\sigma_{\text{edge}}^2} \tag{9}$$

Including all coefficients, related to edges and to noise, the overall distribution of the coefficients $Wf[n,m]$ is given by:

$$p(x) = w_{\text{noise}} p(x|\text{noise}) + (1 - w_{\text{noise}}) p(x|\text{edge}), \tag{10}$$

where w_{noise} is the *a priori* probability for the noise-related coefficient distribution (and, consequently, $1 - w_{\text{noise}}$ is the *a priori* probability for edge-related coefficients).

The parameters σ_{noise}, σ_{edge} and w_{noise} are estimated by maximizing the likelihood function:

$$\ln L = \sum_{(m,n) \text{ in image}} \ln\left(p(Wf[n,m])\right), \tag{11}$$

with the restriction $0 \leq w_{\text{noise}} \leq 1$, where $p(Wf[n,m])$ is the function defined in Equation (10) evaluated at the coefficients $Wf[n,m]$.

Typically, the variance of noise-related coefficient is smaller than the variance of edge-related coefficients. Also, we do not allow the standard deviation of noise-related coefficients to increase as 2^j increases, since noise is filtered out as 2^j increases. Therefore, the restrictions $\sigma_{\text{noise}} < \sigma_{\text{edge}}$ and $\sigma_{\text{noise}}^j \leq \sigma_{\text{noise}}^{j-1}$ are included in the maximization process.

Figure 1(a) shows the original *Lenna* image (512×512 pixels), and Figure 1(b) shows a noisy version of the *Lenna* image (PSNR[1] = 20.17dB). In Figure 2, the histogram of the coefficient $W_{2^2}^1 f[n,m]$ (corresponding to the horizontal detail at the level 2^2) is shown (continuous curve), along with our model fitted to these data (dotted curve). It can be noticed that our model is very close to the actual data.

Once the parameters σ_{noise}, σ_{edge} and w_{noise} are estimated, the conditional probability density functions for the coefficients distributions $p(x|\text{noise})$ and $p(x|\text{edge})$ are given, respectively, by Equations (8) and (9). Also, the *a priori* probabilities for noise-related (w_{noise}) and edge-related ($1 - w_{\text{noise}}$) coefficient distributions are determined by the maximization procedure. The shrinkage function $g(x)$ at each scale and subband is given by the posterior probability function $p(\text{edge}|x)$, which is calculated using Bayes theorem as follows:

$$g(x) = p(\text{edge}|x) = \frac{(1 - w_{\text{noise}}) p(x|\text{edge})}{(1 - w_{\text{noise}}) p(x|\text{edge}) + w_{\text{noise}} p(x|\text{noise})} \tag{12}$$

3.2 Consistency Along Scales

The analysis of the distribution of coefficients $Wf[n,m]$ dos not allow to discriminate between coefficients related to noise from those related to edges. Particularly, coefficients at the level 2^1 are often highly contaminated by noise, and it is difficult to distinguish the distributions associated to noise and edge. For a better discrimination, we combine the shrinkage factors $g_j[n,m]$ along consecutive scales.

At each scale 2^j, the value $g_j[n,m]$ may be interpreted as a confidence measure that the coefficient $W_{2^j} f[n,m]$

[1] The peak-to-peak signal-to-noise ratio (PSNR) is defined by $20 \log_{10}(255/\sigma_{\text{error}})$, where σ_{error} is the standard deviation of the error image $I_{\text{orig}} - I_{\text{noisy}}$.

Figure 1: (a) Original *Lenna* image. (b) Noisy *Lenna* image (PSNR = 20.17 dB).

is in fact associated to an edge. If the value $g_j[n,m]$ is close to 1 for several consecutive levels 2^j, it is more likely that $W_{2^j}f[n,m]$ is associated to an edge. On the other hand, if $g_j[n,m]$ decreases as j increases, it is more likely that $W_{2^j}f[n,m]$ is actually associated to noise. Thus, for the scale 2^j, we combine the shrinkage factors $g_j[n,m]$ in consecutive scales, obtaining the updated shrinkage factor $g_j^{\text{scale}}[n,m]$:

$$g_j^{\text{scale}}[n,m] = \left(\frac{g_j[n,m]^\gamma + \cdots + g_{j+K}[n,m]^\gamma}{K+1} \right)^{1/\gamma}, \tag{13}$$

where γ is an adjustable parameter, and $K+1$ is the number of consecutive scales under consideration. Please, notice that when $\gamma = 1$, the above function is exactly the average of the shrinkage functions. For $\gamma < 1$, smaller coefficients carry more weight, and tend to dominate the summation.

This updating rule is applied from coarser to finer resolution. The shrinkage factor $g_J^{\text{scale}}[n,m]$, corresponding to the coarsest resolution 2^J, is defined as $g_J[n,m]$. However, for other resolutions 2^j, $j = 1...J-1$, the shrinkage factors $g_j^{\text{scale}}[n,m]$ depend on scales $2^j, 2^{j+1}, ..., 2^\kappa$, where $\kappa = \min\{J, j+K\}$. The coefficients $W_{2^j}f$ are then modified according to Equation (7), using the updated shrinkage factors $g_j^{\text{scale}}[n,m]$ instead of $g_j[n,m]$.

3.3 Geometric Consistency

At this point, we have obtained the updated shrinkage functions $g_j^{\text{scale}}[n,m]$, for each level 2^j. However, we may achieve even better discrimination between noise and real edges by imposing geometrical constraints.

Usually, edges appear in contour lines, and not isolated. In our approach, a coefficient $Wf[n,m]$ should have a higher shrinkage factor if its neighbors along the local contour direction also have large shrinkage factors. To enhance the shrinkage factors along contours, we first quantize the gradient directions $\theta_{2^j}f[n,m]$ into $0°, 45°, 90°$ and $135°$. The contour lines are orthogonal to the gradient direction at each edge element, so we can estimate the contour direction from $\theta_{2^j}f[n,m]$. We then add up the shrinkage factors $g_j^{\text{scale}}[n,m]$ along the contour direction, according to the following updating rule:

$$g_j^{\text{geom}}[n,m] = \begin{cases} \sum_{i=-N}^{N} \alpha[i] g_j^{\text{scale}}[n+i,m] \\ \text{if } C_{2^j}[n,m] = 0°, \\ \sum_{i=-N}^{N} \alpha[i] g_j^{\text{scale}}[n+i,m+i] \\ \text{if } C_{2^j}[n,m] = 45°, \\ \sum_{i=-N}^{N} \alpha[i] g_j^{\text{scale}}[n,m+i] \\ \text{if } C_{2^j}[n,m] = 90°, \\ \sum_{i=-N}^{N} \alpha[i] g_j^{\text{scale}}[n+i,m-i] \\ \text{if } C_{2^j}[n,m] = 135°, \end{cases} \tag{14}$$

where $C_{2^j}[n,m]$ is the local contour direction at the pixel $[n,m]$, $2N+1$ is the number of adjacent pixels that should be aligned for geometric continuity, and $\alpha[i]$ is a window that allows neighboring pixels to be weighted differently, according to their distance from the pixel $[n,m]$ under consideration (we used a Gaussian weighting function).

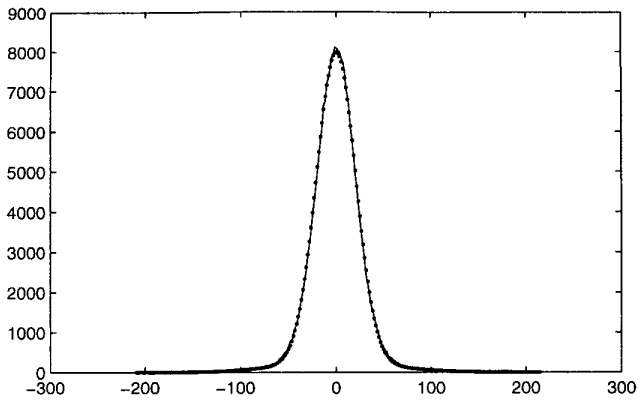

Figure 2: Continuous curve: histogram of the coefficients $W_{2^2}^1 f[n,m]$. Dotted curve: our model fitted to the data.

After updating the shrinkage factors, coefficients with large $g_j^{\text{geom}}[n,m]$ along the local contour direction will be strengthened, while pixels with no geometric continuity will have their shrinkage factors de-enhanced.

4 Experimental Results

We performed a series of denoising experiments to test the performance of our method, using images contaminated with controlled amounts of noise and images with inherent noise. The denoised images were analyzed from the qualitative (visual) and quantitative (PSNR of the filtered image) points of view. For all the examples, we used $\gamma = 0.3$ (Equation (13)), $N = 1$ (Equation (14)), and three dyadic scales in the wavelet decomposition.

Figure 3(a) shows the noisy *Lenna* image filtered using the MATLAB function *wiener2*, which is an implementation of a pixel-wise adaptive Wiener estimate using local statistics. Figure 3(b) shows the same noisy image filtered by our technique. It can be noticed that our method appears to produce a less noisy image, while the edges were kept sharp.

Quantitatively, we compared the performance of our method with four other well established denoising techniques. These methods are: SAWT using an over-complete representation [7], 5×5 LAWMAP [8], the function `wiener2`, and Donoho's hard threshold [15]. Table 1 shows the performance of our technique and these four denoising methods for the *Lenna* image with different amounts of noise. Our method's performance is similar to 5×5 LAWMAP, slightly inferior to SAWT, and much better than `wiener2` and Donoho's hard threshold.

The proposed method was also tested in images with natural noise. Figure 4(a) shows an aerial image (*chemplant* image, 256×256 pixels). Figure 4(b) shows the result of the proposed method. Most of the background noise seems to be removed, while small structures were preserved (buildings, roads, etc.). However, some fine textures (e.g. the crop on the top right) were also smoothed, since it is difficult for our method to distinguish noise from fine textures, as for most methods proposed in the literature.

Noisy *Lenna* image	24.60	22.13	20.17
Hard threshold	28.52	27.24	26.34
`wiener2`	31.12	29.97	28.71
5×5 LAWMAP	32.36	31.01	29.98
SAWT	32.99	31.69	30.61
Our method	31.97	30.93	30.02

Table 1: PSNR results (in dB) of several denoising methods for the *Lenna* image contaminated by different amounts of noise.

We also applied our method to medical images. Figure 5(a) shows a magnetic resonance image (MRI) of a brain (242×248 pixels), and Figure 5(b) shows the filtered image using our technique. It can be noticed that background noise was smoothed out, at the same time that fine small structures were kept.

Our method was implemented in MATLAB, and typical execution time for denoising a 512×512 image in a 350 MHz Pentium II personal computer is about 55 seconds. We believe that implementing our method in a compiled language would improve significantly the execution time.

These experiments show that the proposed technique works well for images contaminated by different amounts of noise. Quantitatively, the method produce PSNR outputs comparable to other state-of-the-art techniques reported in the literature, and qualitatively the denoised images pro-

Figure 3: Noisy *Lenna* image filtered by: (a) *wiener2* (PSNR = 28.71 dB); (b) our technique (PSNR = 30.02 dB).

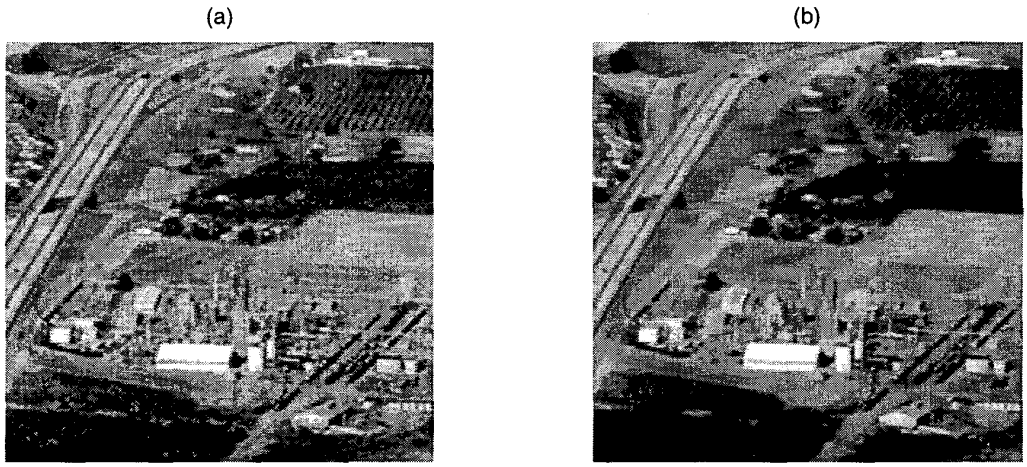

Figure 4: (a) *Chemplant* image. (b) Denoising by our technique.

duce good visual results (noise is reduced, while edges are kept sharp), with the advantages of having comparatively low computational complexity, and being adaptive to different degrees of noise corruption.

5 Concluding Remarks

A new method for image denoising based on the wavelet transform was presented. The experimental results showed that our method produced good quantitative and qualitative results in comparison to other state-of-the-art techniques. Also, the proposed method is adaptive to the amount of noise in the image, performing well for images with inherent noise and images contaminated with artificial noise in different amounts. In this work, the same parameters γ and N were used for all the images. However, these parameters could be fine-tuned to each individual image to produce optimal results. Future work will concentrate on improving our model for the wavelet coefficients, and extending our work to enhancement of noisy images.

References

[1] S. G. Mallat and W. L. Hwang, "Singularity detection and processing with wavelets," *IEEE Transactions on Information Theory*, vol. 38, no. 2, pp. 617–643, 1992.

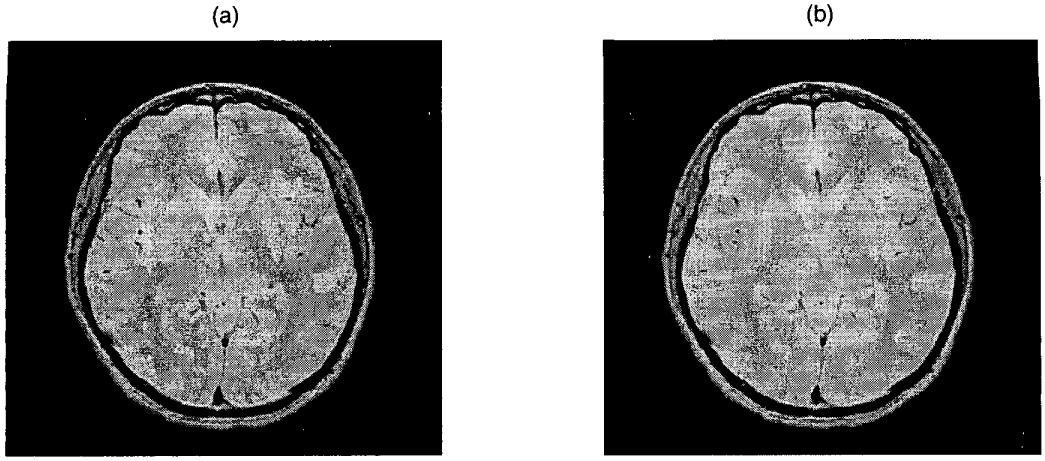

Figure 5: (a) Original MRI image. (b) MRI image filtered by our technique.

[2] R. A. Carmona, "Extrema reconstructions and spline smoothing: Variations on an algorithm of Mallat and Zhong," *Wavelets and Statistics - Lecture Notes in Statistics*, vol. 103, pp. 83–94, 1995.

[3] Y. Xu, J. B. Weaver, D. M. Healy, and J. Lu, "Wavelet transform domain filters: A spatially selective noise filtration technique," *IEEE Transactions on Image Processing*, vol. 3, no. 6, pp. 747–758, 1994.

[4] M. Malfait and D. Roose, "Wavelet based image denoising using a Markov Random Field a priori model," *IEEE Transactions on Image Processing*, vol. 6, no. 4, pp. 549–565, 1997.

[5] E. P. Simoncelli and E. Adelson, "Noise removal via bayesian wavelet coring," in *IEEE International Conference on Image Processing*, (Lausanne, Switzerland), pp. 279–382, September 1996.

[6] S. G. Chang and M. Vetterli, "Spatial adaptive wavelet thresholding for image denoising," in *1997 International Conference on Image Processing (ICIP97)*, pp. 374–377, 1997.

[7] S. G. Chang, B. Yu, and M. Vetterli, "Spatially adaptive wavelet thresholding with context modeling for image denoising," *IEEE Transactions on Image Processing*, vol. 9, pp. 1522–1531, September 2000.

[8] M. K. Mihçak, I. Kozintsev, K. Ramchandram, and P. Moulin, "Low-complexity image denoising based on statistical modeling of wavelet coefficients," *IEEE Signal Processing Letters*, vol. 6, pp. 300–303, December 1999.

[9] V. Strela, J. Portilla, and E. P. Simoncelli, "Image denoising via a local gaussian scale mixture model in the wavelet domain," in *SPIE 45th Annual Meeting*, (San Diego, CA), August 2000.

[10] A. Pizurica, W. Philips, I. Lemahieu, and M. Acheroy, "Image de-noising in the wavelet domain using prior spatial constraints," in *Seventh International Conference on Image Processing and its Applications*, (Manchester, England), pp. 216–219, IEE, July 1999.

[11] S. G. Mallat and S. Zhong, "Characterization of signals from multiscale edges," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 7, pp. 710–732, 1992.

[12] S. G. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 2, no. 7, pp. 674–693, 1989.

[13] D. L. Donoho, "Nonlinear wavelet methods for recovery of signals, densities and spectra from indirect and noisy data," in *Proc Symp Appl Math* (I. Daubechies, ed.), (Providence, RI), 1993.

[14] D. L. Donoho, "Wavelet shrinkage and w.v.d.: A 10-minute tour," *Progress in Wavelet Analysis and Applications*, 1993.

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