Adaptive Image Denoising Using Scale and Space Consistency

Jacob Scharcanski, Cláudio R. Jung, and Robin T. Clarke

***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, whilst image noise is filtered out. At each resolution level, the image edges are estimated by gradient magnitudes (obtained from the wavelet coefficients), which are modeled probabilistically, and a shrinkage function is assembled based on the model obtained. Joint use of space and scale consistency is applied for better preservation of edges. The shrinkage functions are combined to preserve edges that appear simultaneously at several resolutions, and geometric constraints are applied to preserve edges that are not isolated. The proposed technique produces a filtered version of the original image, where homogeneous regions appear separated by well-defined edges. Possible applications include image presegmentation, and image denoising.**

***Index Terms*—Edge detection, image denoising, multiresolution analysis, wavelets.**

I. Introduction

IN IMAGE analysis, removal of noise without blurring the image edges is a difficult problem. Typically, noise is characterized by high spatial frequencies in an image, and Fourier-based methods usually try to suppress high-frequency components, which also tend to reduce edge sharpness.

On the other hand, the wavelet transform provides good localization in both spatial and spectral domains, and low-pass filtering is inherent to this transform. There are now several approaches for noise suppression using wavelets, which have shown 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 (more exactly, only the local maxima are used). The reconstruction process is based on an interactive projection procedure, which may be computationally demanding.

Lu *et al.* [2] have proposed using wavelets for image filtering and edge detection. In their approach, local maxima are tracked in scale-space, and represented by a tree structure. A metric is applied to prune the tree, removing local maxima related to false edges. Finally, the inverse wavelet transform is applied, and the output is the denoised image with edge preservation. However, construction of the tree is difficult for noisy images containing edges of various local contrasts (there are erroneous connections when the wavelet coefficient maxima are dense). In this case, some edges are lost, and filtering may not be efficient. Other denoising methods based on wavelet coefficient trees were proposed by Donoho [3] and Baraniuk [4].

Xu *et al.* [5] 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 [6] developed a filtering technique that takes into account two measures for image filtering. The first is a measure of local regularity of the image through 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 signal-to-noise ratio (SNR) gain achieved by this method is significant, but the stochastic sampling procedure needed for the probabilities calculation is computationally demanding. Another approach that uses a Markov random field model for wavelet-based image denoising was proposed by Jansen and Bulthel [7].

Other authors also proposed probabilistic approaches for image denoising in the wavelet domain. Simoncelli and Adelson [8] used a two-parameter generalized Laplacian distribution for the wavelet coefficients of the image, which is estimated from the noisy observations. Chang *et al.* [9] proposed the use of adaptive wavelet thresholding for image denoising, by modeling the wavelet coefficients as a generalized Gaussian random variable, whose parameters are estimated locally (i.e., within a given neighborhood). Strela *et al.* [10] 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. All these methods mentioned above require a

Manuscript received September 18, 2000; revised May 3, 2002. This work was supported by the Fundação de Amparo a Pesquisa do Estado do Rio Grande do Sul, Brazil, (FAPERGS) and the Conselho Nacional de Desenvolvimento Cientifico e Tecnológico, Brazil (CNPq). The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Uday B. Desai.

J. Scharcanski is with the Instituto de Informática, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brazil 91501-970 (e-mail: jacobs@inf.ufrgs.br).

C. R. Jung was with the Instituto de Informática, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brazil 91501-970. He is now with the Centro de Ciências Exatas e Tecnológicas, Universidade do Vale do Rio dos Sinos, São Leopoldo, RS, Brazil 93022-000 (e-mail: crjung@exatas.unisinos.br).

R. T. Clarke is with Instituto de Pesquisas Hidráulicas, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brazil 91501-970 (e-mail: clarke@iph.ufrgs.br).

Publisher Item Identifier 10.1109/TIP.2002.802528.

noise estimate, which may be difficult to obtain in practical applications.

Pizurica *et al.* [11] 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 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.

This paper proposes a new method for image denoising using the wavelet transform, which combines wavelet coring and the joint use of scale and space consistency. The image gradient is calculated from the detail images (horizontal and vertical) of the wavelet transform, and the distribution of the gradient magnitudes associated to edges and noise are modeled by Rayleigh probability density functions. A *shrinkage function*, assuming values between zero and one, 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 IV presents some experimental results for our approach, and a comparison with other denoising techniques. Conclusions are presented in the final section.

II. Wavelet Transform in Two Dimensions

In this work, the two-dimensional (2-D) wavelet decomposition uses only two detail images (horizontal and vertical details) [12], instead of the already conventional approach in which three detail images (horizontal, vertical, and diagonal details) are used [13]. This 2-D wavelet transform requires two wavelets, namely, $\psi^1(x, y)$ and $\psi^2(x, y)$. At a particular scale s we have

$$\psi_s^i(x, y) = \frac{1}{s^2}\,\psi^i\left(\frac{x}{s}, \frac{y}{s}\right), \qquad i = 1, 2. \tag{1}$$

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

$$W_{2^j}^i f(x, y) = \left(f * \psi_{2^j}^i\right)(x, y), \qquad i = 1, 2. \tag{2}$$

Therefore, the multiresolution wavelet coefficients are

$$\mathbf{W}_{2^j} f(x, y) = \left(W_{2^j}^1 f(x, y), W_{2^j}^2 f(x, y)\right). \tag{3}$$

The original signal $f(x, y)$ is then represented by the 2-D wavelet transform, in terms of the two dual wavelets $\xi^1(x, y)$ and $\xi^2(x, y)$

$$f(x, y) = \sum_j \left(\left(W_{2^j}^1 f * \xi_{2^j}^1\right)(x, y) + \left(W_{2^j}^2 f * \xi_{2^j}^2\right)(x, y)\right). \tag{4}$$

In order to build a multiscale representation, we need a scaling function $\phi(x, y)$ (which is a low-pass filter), and the corresponding component at a scale 2^j is

$$S_{2^j} f(x, y) = (f * \phi_{2^j})(x, y). \tag{5}$$

We may interpret the component $S_{2^J} f(x, y)$ as a smoothed version of $f(x, y)$, and the components $\mathbf{W}_{2^j} f(x, y)$, for $j = 1, \ldots, J$, as the image details lost by smoothing going from $S_{2^0} f(x, y)$ to $S_{2^J} f(x, y)$. Further details may be found in [12] and [13].

A. Edge Detection Using Wavelets

Now, it is necessary to find a wavelet basis such that its components $\mathbf{W}_{2^j} f(x, y)$ are related to the local gradients of the image at the scale 2^{j-1}. A smoothing function $\theta(x, y)$ [which is different from the scaling function $\phi(x, y)$, and used only to define the wavelets $\psi^1(x, y)$ and $\psi^2(x, y)$] is selected, and the wavelets are defined as

$$\psi^1(x, y) = \frac{\partial}{\partial x}\,\theta(x, y) \quad \text{and} \quad \psi^2(x, y) = \frac{\partial}{\partial y}\,\theta(x, y). \tag{6}$$

Note that the wavelet coefficient $\mathbf{W}_s f(x, y)$ can be written as

$$\begin{aligned}\mathbf{W}_s f(x, y) &= \begin{pmatrix} W_s^1 f(x, y) \\ W_s^2 f(x, y)\end{pmatrix} \\ &= s\begin{pmatrix} \dfrac{\partial}{\partial x}(f * \theta_s)(x, y) \\ \dfrac{\partial}{\partial y}(f * \theta_s)(x, y)\end{pmatrix} \\ &= s\nabla(f * \theta_s)(x, y)\end{aligned} \tag{7}$$

which in fact corresponds to the gradient of the smoothed version of f at the scale s. Observing that an edge can be defined as a local maximum of the gradient modulus along the gradient direction [14], we can detect the edges at the scale s from $\mathbf{W}_s f(x, y)$. A suitable choice for $\theta(x, y)$ proposed in [12] was a cubic spline with compact support. This approach can be used for digital images $f[n, m]$, using a discrete version of the wavelet transform [12].

III. Our Image Denoising Approach

Given a digital image $f[n, m]$, we first apply the redundant wavelet transform using only two detail images, as discussed in the previous section. As a result, at each resolution 2^j, we obtain the detail images $W_{2^j}^1 f$, $W_{2^j}^2 f$ and the smoothed image $S_{2^j} f$. The edge magnitudes can be calculated from the image gradient, as follows:

$$M_{2^j} f = \sqrt{\left(W_{2^j}^1 f\right)^2 + \left(W_{2^j}^2 f\right)^2} \tag{8}$$

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

$$\varsigma_{2^j} f = \arctan\left(\frac{W_{2^j}^2 f}{W_{2^j}^1 f}\right). \tag{9}$$

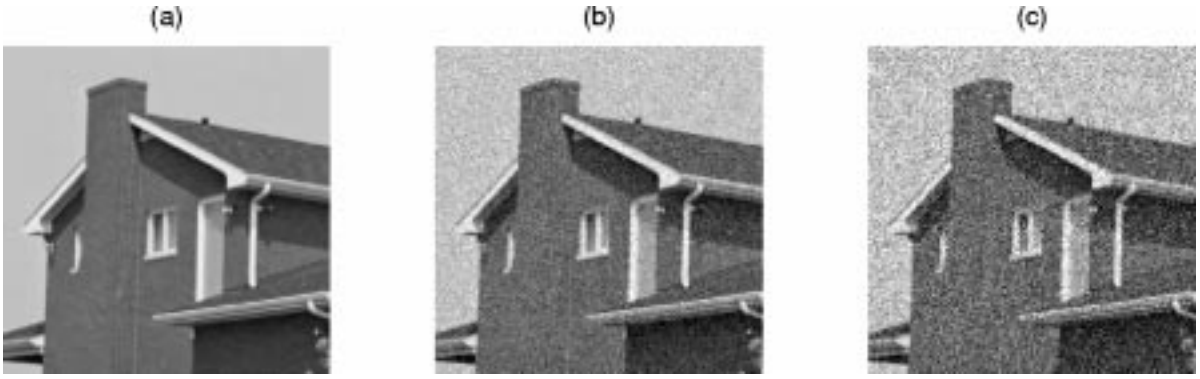

Fig. 1. (a) Original *house* image. (b) First noisy *house* image (SNR = 8 dB). (c) Second noisy *house* image (SNR = 3 dB).

Due to noise, some pixels of homogeneous regions may have gradient magnitudes $M_{2^j}f$ that could be misinterpreted as edges, so we next 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.

A. Wavelet Coring

Image coring is a known approach for noise reduction, where the image highpass bands are subject to a nonlinearity that reduces (or suppresses) low-amplitude values and retains high-amplitude values [8]. Many variants of coring have been developed, and the concept of "shrinkage" has been used with wavelets [15].

For each level 2^j, we want to find a nonnegative nondecreasing shrinkage function $g_j(x)$, $0 \leq g_j(x) \leq 1$, such that the wavelet coefficients $W^1_{2^j}f$ and $W^2_{2^j}f$ are updated according the following rule:

$$NW^i_{2^j}f[n, m] = W^i_{2^j}f[n, m]g_j(M_{2^j}f[n, m]), \qquad \text{for } i = 1, 2 \quad (10)$$

where $g_j(M_{2^j}f[n, m])$ is a *shrinkage factor*. Write $g_j[n, m] = g_j(M_{2^j}f[n, m])$. To find the functions g_j, we analyze the magnitude image $M_{2^j}f$. Some of these coefficients are related to noise, and others to edges. If the image is contaminated by additive white noise, the corresponding coefficients $W^1_{2^j}f$ and $W^2_{2^j}f$ may be considered Gaussian distributed [16], with standard deviation σ^j_{noise}. As a consequence, the distribution of the corresponding magnitudes $M_{2^j}f = \sqrt{(W^1_{2^j}f)^2 + (W^2_{2^j}f)^2}$, at each resolution 2^j, may be approximated by a Rayleigh probability density function [17]

$$p_j(r|\text{noise}) = \frac{r}{[\sigma^j_{\text{noise}}]^2}\, e^{-r^2/[2\sigma^j_{\text{noise}}]^2}. \quad (11)$$

However, in practice, we observe that noise-free images typically consist of homogeneous regions and not many edges. In general, homogeneous regions contribute with a sharp peak around zero for the histograms of $W^1_{2^j}f$ and $W^2_{2^j}f$, and the edges contribute to the tail of the distribution. This distribution presents a sharper peak than a Gaussian [8], and therefore, the Gaussian model is not appropriate for the distribution of the coefficients. In fact, other distributions have been used for modeling the wavelet coefficients, such as two-parameter generalized Laplacian distributions [8], Gaussian distributions with high local correlation [18], generalized Gaussian distributions [9] and Gaussian mixtures [10], [19]. However, we assume that the distribution of the wavelet coefficients $W^1_{2^j}f$ and $W^2_{2^j}f$ related exclusively to edges (and not related to homogeneous regions) is approximated by a Gaussian (i.e., when the sharp peak in $W^1_{2^j}f$ and $W^2_{2^j}f$ associated to homogeneous regions is not considered, we assume that the remaining data is approximated by a Gaussian). The normal model for edge-related coefficients is assumed because it leads to a simple model (Rayleigh) to approximate the corresponding edge-related gradient magnitudes $M_{2^j}f$.

For example, consider the 256 × 256 *house* image and its noisy versions (SNR = 8 dB and 3 dB), shown in Fig. 1, from left to right. Fig. 2 shows normal plots of the coefficients $W^1_{2^1}f$ for the *house* images (corresponding to the finest resolution of the horizontal subband). In Fig. 2(a), all the coefficients $W^1_{2^1}f$ for the original *house* image were used. This distribution shows significant departure from a Gaussian distribution, as expected [8]. However, Fig. 2(b), showing the Normal plot obtained using only the edge-related coefficients from the original *house* image, shows an acceptable agreement with the linearity expected under the Gaussian hypothesis, even considering that coefficients associated exclusively to edges are difficult to isolate in experiments. We conclude that the Gaussian assumption for edge-related coefficients is not unreasonable. Finally, Fig. 2(c) corresponds to the first noisy *house* image (SNR = 8 dB), and an even closer match to the Gaussian distribution is noticed. This match occurs because noise typically affects all the wavelet coefficients in the image, while edges are related to just few image coefficients (and thus the noise distribution dominates over the edge distribution).

Therefore, the edge-related magnitudes $M_{2^j}f$ are approximated by a Rayleigh process

$$p_j(r|\text{edge}) = \frac{r}{[\sigma^j_{\text{edge}}]^2}\, e^{-r^2/2[\sigma^j_{\text{edge}}]^2}. \quad (12)$$

The overall gradient magnitude distribution $M_{2^j}f$ (including coefficients related to edges and noise) is given by

$$p_j(r) = w^j_{\text{noise}}p_j(r|\text{noise}) + (1 - w^j_{\text{noise}})p_j(r|\text{edge}) \quad (13)$$

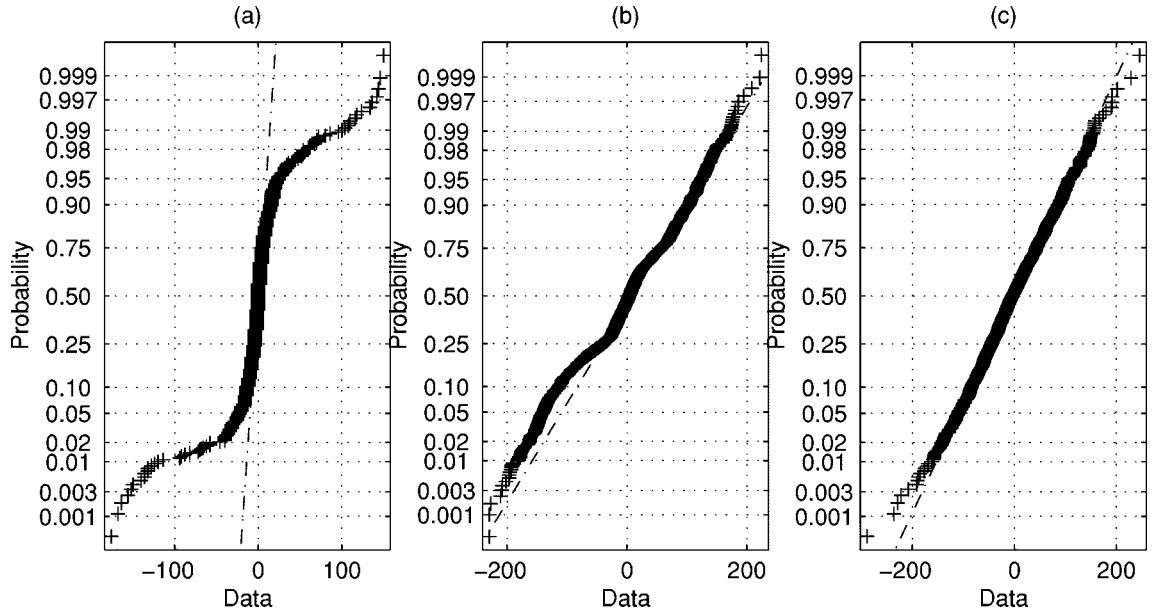

Fig. 2. Normal plots of the coefficients $W^1_{2^j} f[n, m]$. (a) Using all coefficients for the original *house* image. (b) Using only edge-related coefficients for the original *house* image. (c) Using all the coefficients for the first noisy *house* image (SNR = 8 dB).

where w^j_{noise} is the *a priori* probability for the noise-related gradient magnitude distribution (and, consequently, $1 - w^j_{\text{noise}}$ is the *a priori* probability for edge-related gradient magnitudes). To simplify the notation, we remove the index j, and (13) can thus be written as

$$p(r) = w_{\text{noise}} p(r|\text{noise}) + (1 - w_{\text{noise}}) p(r|\text{edge}). \tag{14}$$

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

$$\ln L = \sum_{(m,n)\text{ in image}} \ln(p(M_{2^j} f[m, n])) \tag{15}$$

with the restriction $0 \leq w_{\text{noise}} \leq 1$, where $p(M_{2^j} f[m, n])$ is the function defined in (14) evaluated at the gradient magnitudes $M_{2^j} f[m, n]$.

Typically, the number of noise-related coefficients is much larger than those related to edges [as suggested by Fig. 2(c)], and also their magnitudes are usually smaller. Therefore, the peak of the gradient magnitude histogram is mostly due to noise-related coefficients, and usually is approximately at the same location as the peak of the noise-related magnitude distribution $p(r|\text{noise})$. Considering that the mode of the Rayleigh probability density function $p(r|\text{noise})$ is given by σ_{noise} [17], we can estimate the parameter σ_{noise} as the localization of the magnitude histogram peak. The computational cost involved in the maximization of (15) is then reduced, because only two parameters (w_{noise} and σ_{edge}) are utilized, given the restriction $\sigma_{\text{noise}} < \sigma_{\text{edge}}$. This procedure is adaptive and does not require a noise estimate.

Fig. 3 shows histograms of gradient magnitudes for the 8-dB and 3-dB noisy versions of the *house* image, and the obtained model $p(r)$ for these distributions, at the resolutions 2^1, 2^2, and 2^3. It is seen that the histograms are well approximated by our model, and no further noise estimates are needed.

Once the parameters σ_{noise}, σ_{edge} and w_{noise} are estimated, the conditional probability density functions for the gradient magnitude distributions $p(r|\text{noise})$ and $p(r|\text{edge})$ are given, respectively, by (11) and (12). Also, we have determined the *a priori* probabilities for noise-related (w_{noise}) and edge-related ($1 - w_{\text{noise}}$) gradient magnitude distributions. The shrinkage function $g_j(x)$ for each resolution 2^j is given by the posterior probability function $p(\text{edge}|r)$, which is calculated using Bayes theorem as follows:

$$p(\text{edge}|r) = \frac{(1 - w_{\text{noise}}) p(r|\text{edge})}{(1 - w_{\text{noise}}) p(r|\text{edge}) + w_{\text{noise}} p(r|\text{noise})}. \tag{16}$$

For the second noisy *house* image, the spatial occurrence of the shrinkage factors $g_j[n, m]$, for $j = 1, 2, 3$, are shown in Fig. 4. Brighter pixels correspond to factors close to one, while darker pixels correspond to factors close to zero. At the finest resolution (2^1), noise-related and edge-related coefficients have almost the same magnitude. As a consequence, the discrimination between edge- and noise-related coefficients is difficult, as seen in Fig. 4(a). For the lower resolution levels (2^2 and 2^3), the results are more reliable, since noise is smoothed out when the resolution decreases (j increases). Further discrimination can be achieved by analyzing the evolution of the shrinkage functions along consecutive scales and applying spatial constraints, as discussed in the next section.

B. Scale and Spatial Constraints

1) Consistency Along Scales: It is known that coefficients associated with noise tend to vanish as the level 2^j increases, while coefficients associated with edges tend to be preserved

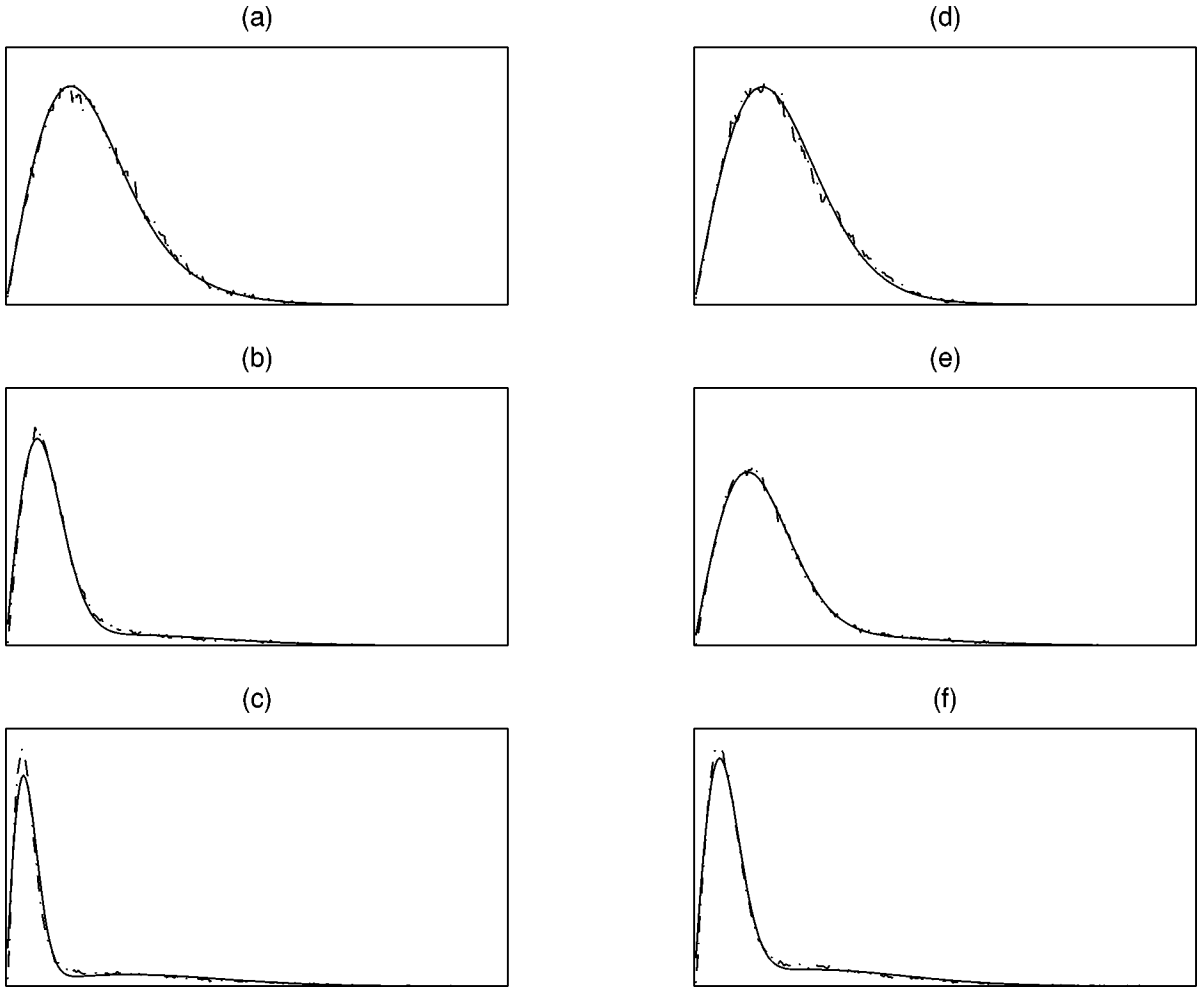

Fig. 3. (a)–(c) Histogram of the gradient magnitudes (dash-dotted line) and the estimated magnitude density function (solid line) for the first noisy *house* image (SNR = 8 dB), at the resolutions 2^1, 2^2 and 2^3. (e)–(f) Same as (a)–(c), but for the second noisy *house* image (SNR = 3 dB).

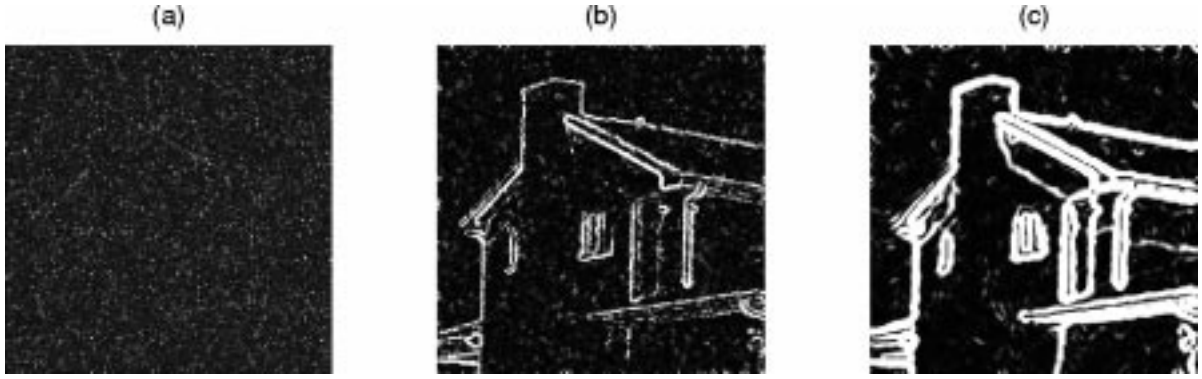

Fig. 4. From left to right: shrinkage factors $g_j[n, m]$, for $j = 1, 2, 3$, for the second noisy *house* image.

when j increases. In [1] and [6] the Hölder exponent was calculated in order to explore this property. We analyze the consistency of the wavelet coefficients along scales (i.e., resolutions) differently, by combining the shrinkage functions g_j at various resolutions 2^j.

For each scale 2^j, the value $g_j[n, m]$ may be interpreted as a confidence measure that the coefficient $M_{2^j}[n, m]$ is in fact associated to an edge. If the value $g_j[n, m]$ is close to unity for several consecutive levels 2^j, it is more likely that $M_{2^j}[n, m]$ is associated with an edge. On the other hand, if $g_j[n, m]$ decreases as j increases, it is more likely that $M_{2^j}[n, m]$ is actually associated with noise.

For each scale 2^j, we use the information provided by the function g_j, and also by the functions g_{j+k}, for $k = 1, 2, \ldots, K$, where $K + 1$ is the number of consecutive resolutions that will be taken into consideration for the consistency along scales. As observed by Xu *et al.* [5], it appears that when two or three consecutive resolutions are used, better

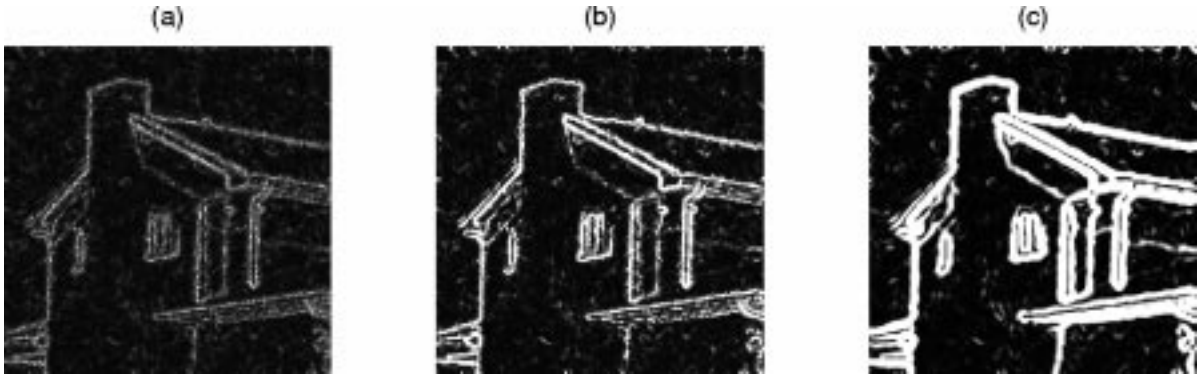

Fig. 5. From left to right: shrinkage factors $g_j^{\text{scale}}[n, m]$, for $j = 1, 2, 3$, after consistency along scales was applied, for the second noisy *house* image.

results are obtained than from using more consecutive resolutions, because the positions of the local maxima of $M_{2^j}[n, m]$ may change as j increases.

Thus, we need to find a function h: $\mathbb{R}^{K+1} \rightarrow \mathbb{R}$ such that $h(x_1, x_2, \ldots, x_{K+1})$ is approximately one if all the x_i are close to one, and $h(x_1, x_2, \ldots, x_{K+1})$ must be close to zero if any of the x_i is close to zero. There are many functions satisfying this property, and we chose the harmonic mean

$$h(x_1, x_2, \ldots, x_{K+1}) = \frac{K+1}{\frac{1}{x_1} + \frac{1}{x_2} + \cdots + \frac{1}{x_{K+1}}}. \quad (17)$$

For the scale 2^j, the updated function g_j^{scale} is given by

$$g_j^{\text{scale}}(x) = \frac{K+1}{\frac{1}{g_j(x)} + \frac{1}{g_{j+1}(x)} + \cdots + \frac{1}{g_{j+K}(x)}}. \quad (18)$$

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 equal to $g_J[n, m]$. However, for other resolutions 2^j, $j = 1 \cdots J-1$, the shrinkage factors $g_j^{\text{scale}}[n, m]$ depend on scales 2^j, 2^{j+1}, $\ldots$, 2^κ, where $\kappa = \min\{J, j+K\}$. The coefficients $W_{2^j}^1 f$ and $W_{2^j}^2 f$ are then modified according to (10), using the updated shrinkage factors $g_j^{\text{scale}}[n, m]$ instead of $g_j[n, m]$. For the second noisy *house* image, the spatial occurrences of the updated shrinkage factors $g_j^{\text{scale}}[n, m]$, for $j = 1, 2, 3$, are shown in Fig. 5.

2) Geometric Consistency: At this point, we have obtained the shrinkage factors $g_j^{\text{scale}}[n, m]$ for each level 2^j. However, we may achieve even better discrimination between noise and edges by imposing geometrical constraints. Usually, edges do not appear isolated in an image. They form contour lines, which we assume to be polygonal (i.e., piecewise linear). In our approach, a coefficient $M_{2^j} f[n, m]$ should have a higher shrinkage factor if its neighbors along the local contour direction also have large shrinkage factors. To detect this kind of behavior, we first quantize the gradient directions $\varsigma_{2^j} f$ into 0°, 45°, 90°, or 135°. The contour lines are orthogonal to the gradient direction at each edge element, so we can estimate the contour direction from $\varsigma_{2^j} f$. We then add up the shrinkage factors $g_j^{\text{scale}}[n, m]$ along the contour direction, according to the following updating rule:

$$g_j^{\text{geom}'}[n, m] = \begin{cases} \displaystyle\sum_{i=-N}^{N} \alpha[i] g_j^{\text{scale}}[n+i, m], \\ \quad \text{if } C_{2^j}[n, m] = 0°, \\ \displaystyle\sum_{i=-N}^{N} \alpha[i] g_j^{\text{scale}}[n+i, m+i], \\ \quad \text{if } C_{2^j}[n, m] = 45°, \\ \displaystyle\sum_{i=-N}^{N} \alpha[i] g_j^{\text{scale}}[n, m+i], \\ \quad \text{if } C_{2^j}[n, m] = 90°, \\ \displaystyle\sum_{i=-N}^{N} \alpha[i] g_j^{\text{scale}}[n+i, m-i], \\ \quad \text{if } C_{2^j}[n, m] = 135° \end{cases} \quad (19)$$

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.

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 not have their shrinkage factors enhanced. This approach has limitations close to corners and junctions, where two or more different local contour directions arise. If the image is sufficiently sampled, high curvature points should also be enhanced.

In the presence of noise, randomly aligned coefficients occur, and could also be strengthened. To overcome this potential difficulty, we compare the contour direction in two consecutive levels. It is expected that contours would be aligned along the same direction in two consecutive levels (it is the same contour at different resolutions), but responses due to noise should not be aligned (gradients associated to noise will not be oriented consistently in consecutive resolutions). Therefore, a second updating rule is applied to the shrinkage factors $g_j^{\text{geom}'}[n, m]$. The

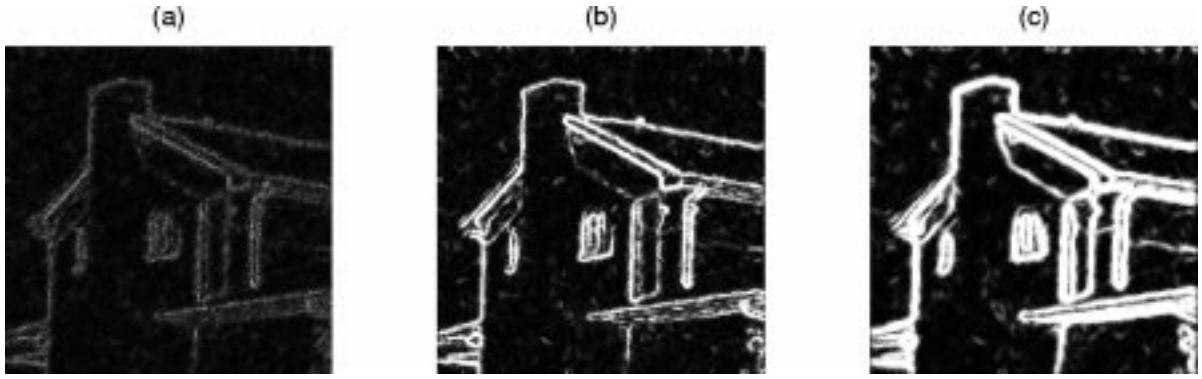

Fig. 6. From left to right: shrinkage factors $g_j^{\text{geom}}[n, m]$, for $j = 1, 2, 3$, after geometric constraints were applied, for the second noisy *house* image.

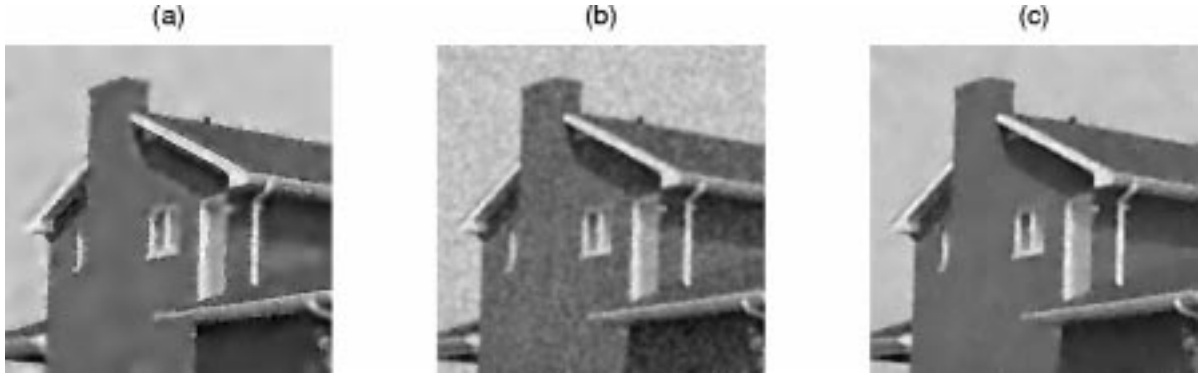

Fig. 7. Results of denoising techniques for the second noisy *house* image. (a) `wave2©` software. (b) Wiener filtering. (c) Our method.

second updating rule takes into account the normalized inner product of corresponding vectors in consecutive resolutions

$$g_j^{\text{geom}}[n, m] = g_j^{\text{geom}'}[n, m] \cdot |\cos(\varsigma_{2^j} f[n, m] - \varsigma_{2^{j+1}} f[n, m])|. \quad (20)$$

Notice that the factor $|\cos(\varsigma_{2^j} f[n, m] - \varsigma_{2^{j+1}} f[n, m])|$ provides a measure for direction continuity. It has value one if the same direction occurs in two consecutive levels, and value zero if the orientations differ by 90° (i.e., are orthogonal). Fig. 6 shows the spatial occurrences of the shrinkage factors $g_j^{\text{geom}}[n, m]$ for the second noisy *house* image, after applying the local geometrical constraints.

C. Overview of the Proposed Method

A schematic overview of our method is as follows.

1) Compute the wavelet transform, obtaining the coefficients $W_{2^j}^1 f$, $W_{2^j}^2 f$ and $S_{2^j} f$.
2) Calculate the edge magnitudes $M_{2^j} f = \sqrt{(W_{2^j}^1 f)^2 + (W_{2^j}^2 f)^2}$ and orientations $\varsigma_{2^j} f = \arctan(W_{2^j}^2 f / W_{2^j}^1 f)$.
3) Compute the parameters σ_{noise}, σ_{edge} and w_{noise}, and then calculate the shrinkage factors $g_j[n, m]$.
4) Combine the shrinkage factors $g_j[n, m]$ in consecutive scales, obtaining the updated shrinkage factors

$$g_j^{\text{scale}}[n, m] = \frac{K+1}{\frac{1}{g_j[n,m]} + \frac{1}{g_{j+1}[n,m]} + \cdots + \frac{1}{g_{j+K}[n,m]}}$$

where $K + 1$ is the number of consecutive scales to be analyzed.
5) Apply a second updating rule to the factors $g_j^{\text{scale}}[n, m]$ using geometrical constraints (contour continuity and orientation continuity along consecutive levels), obtaining $g_j^{\text{geom}}[n, m]$.
6) Modify the coefficients $W_{2^j}^1 f$ and $W_{2^j}^2 f$, obtaining $NW_{2^j}^i f[n, m] = W_{2^j}^i f[n, m] g_j^{\text{geom}}[n, m]$, for $i = 1, 2$.
7) Apply the inverse wavelet transform with $S_{2^J} f$ and the updated coefficients $NW_{2^j}^1 f$ and $NW_{2^j}^2 f$, obtaining the filtered image.

IV. Experimental Results

We applied our technique to images with natural and artificial noise, and compared the results with those obtained by two denoising methods. The first is the `wiener2` function implemented in MATLAB, based on 2-D Wiener filtering [20]. The second is the software `wave2©`, which is an implementation of the method described in [1]. The chosen parameter values for

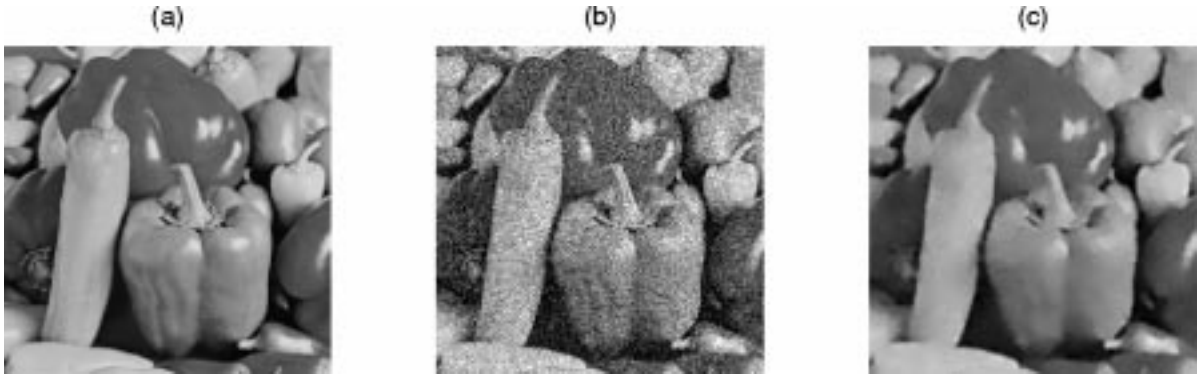

Fig. 8. (a) Original *peppers* image. (b) Noisy *peppers* image (SNR = 3 dB). (c) Result of our method.

the software wave2© are the same as those used by Malfait and Roose [6]. To evaluate the performance of the method, both visual quality and SNR gain are utilized.

In our experiments, the value $N = 3$ was used for geometric continuity in (19), and the window $\alpha[i]$ is a Gaussian (so that larger weights are assigned to the nearest neighbors). A reliable estimate for the parameter σ_{noise} is obtained by first smoothing the magnitude histogram with a Gaussian, and then finding the localization of the peak.

Fig. 7 shows the results of the software wave2© applied to the 3-dB noisy *house* image, followed by the results that were obtained applying the standard Wiener filtering and our technique. Quantitatively, filtering by software wave2© resulted in a SNR of 12.83 dB, by Wiener filtering the output image had SNR = 12.63 dB, and filtering with our technique resulted in a SNR of 15 dB. A similar image (noisy *house* image also with SNR = 3 dB) was used in [6], and the resulting filtered image achieved SNR = 14.86 dB. Qualitatively, it is possible to see that the output of our technique is both sharper and less noisy than the other two methods.

The *peppers* image was also used to test the performance of our method. Fig. 8 shows the original *peppers* image on the left. The middle image is the noisy version (SNR = 3 dB), and the image on the right is the result of our method (SNR = 13.81 dB). All images have a resolution of 256 × 256 pixels. The image processed with the new technique has a visually acceptable quality, and the SNR gain achieved is considerable. The outputs for the wave2© software and Wiener filtering produced, respectively, outputs with SNR = 11 dB and 12.46 dB. Malfait and Roose [6] also used the *peppers* image with added noise (SNR = 3 dB) in their experiments, and their denoising method achieved SNR = 12.36 dB.

We also used images with inherent natural noise in our experiments. Fig. 9(a) shows a natural aerial scene (250 × 500 pixels), while Fig. 9(b) and (c) show, respectively, the denoised images obtained with our technique and Wiener filtering. Noise is effectively removed by our technique and edges were preserved, although some subtle textures were lost. Visual comparison favors our method in comparison to conventional techniques, such as Wiener filtering. Another aerial image (256 × 256 pixels) is shown in Fig. 10(a), and the output of our method and Wiener filtering are shown in Fig. 10(b) and (c), respectively.

A brain magnetic resonance image (MRI) is shown in Fig. 11(a), and the denoised images corresponding to the proposed method and Wiener filtering are shown, respectively, in Fig. 11(b) and (c). It can be noticed that noise reduction was remarkable in Fig. 11(b), while low-contrast structures were preserved. Also, the edges in Fig. 11(b) appear to be sharper than those in Fig. 11(c).

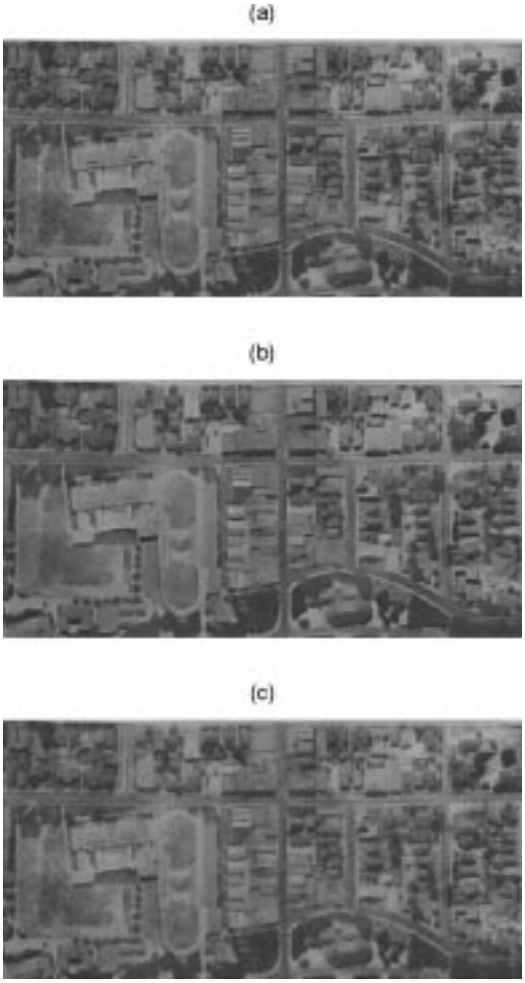

Fig. 9. (a) First aerial image. (b) Filtered image, using our method. (c) Filtered image, using Wiener filtering.

Our algorithm was implemented in MATLAB, running on a 300-MHz Pentium II personal computer, with 64 MB RAM. Typical execution time for a 256 × 256 image, using three dyadic scales, is about 90 s. Most of the running time is dedicated to the

Fig. 10. (a) Second aerial image. (b) Filtered image, using our method. (c) Filtered image, using Wiener filtering.

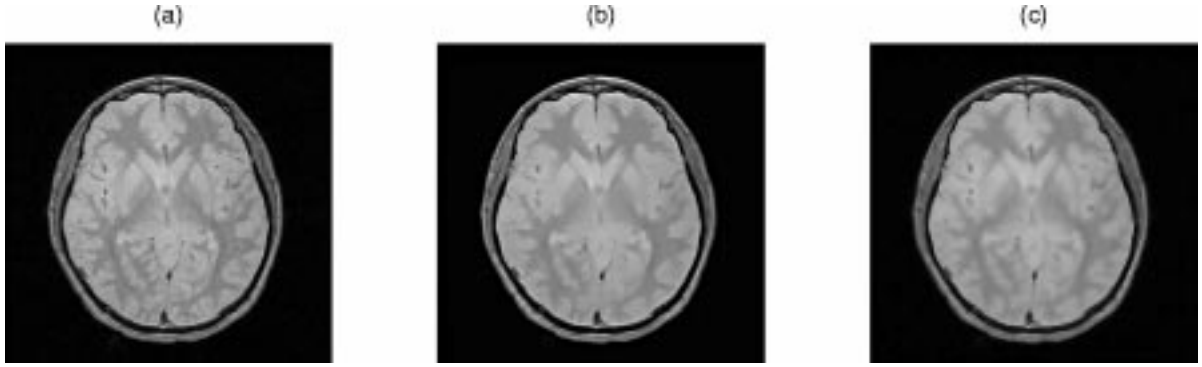

Fig. 11. (a) Original brain MRI. (b) Filtered image, using our method. (c) Filtered image, using Wiener filtering.

maximization of (15). An efficient implementation on a compiled language is expected to improve the execution time.

V. Conclusion

Our denoising procedure consist basically of three steps. Initially, a shrinkage function for each level is assembled modeling the distribution of the gradient magnitudes using Rayleigh probability density functions. Next, scale and spatial constraints are applied. The shrinkage functions are combined in consecutive resolutions, using scale consistency criteria. Finally, geometrical constraints are applied to enhance edges appearing as contours, and therefore connected.

The experimental results obtained are promising, both quantitatively and qualitatively. From this point of view, the new method is comparable to, or better than, other denoising techniques, with the advantage of being adaptive (no estimate of the noise is needed, as opposed to [6], [8]–[10]).

Future work will concentrate on finding more accurate models for the gradient magnitude distribution, distinct choices of shrinkage functions, and a probabilistic approach for multiscale consistency. Also, we intend to investigate the application of our method to edge enhancement in noisy images.

References

[1] S. G. Mallat and W. L. Hwang, "Singularity detection and processing with wavelets," *IEEE Trans. Inform. Theory*, vol. 38, pp. 617–643, Mar. 1992.

[2] J. Lu, J. B. Weaver, D. M. Healy, and Y. Xu, "Noise reduction with multiscale edge representation and perceptual criteria," in *Proc. IEEE-SP Int. Symp. Time-Frequency and Time-Scale Analysis*, Victoria, BC, Oct. 1992, pp. 555–558.

[3] D. L. Donoho, "CART and best-ortho-basis: A connection," *Ann. Statist.*, pp. 1870–1911, 1997.

[4] R. G. Baraniuk, "Optimal tree approximation with wavelets," in *Proc. SPIE Tech. Conf. Wavelet Applications Signal Processing VII*, vol. 3813, Denver, CO, 1999, pp. 196–207.

[5] Y. Xu, J. B. Weaver, D. M. Healy, and J. Lu, "Wavelet transform domain filters: A spatially selective noise filtration technique," *IEEE Trans. Image Processing*, vol. 3, pp. 747–758, Nov. 1994.

[6] 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.

[7] M. Jansen and A. Bulthel, "Empirical bayes approach to improve wavelet thresholding for image noise reduction," *Journal of the American Statistical Association*, vol. 96, no. 454, pp. 629–639, June 2001.

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

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

[10] V. Strela, J. Portilla, and E. P. Simoncelli, "Image denoising via a local Gaussian scale mixture model in the wavelet domain," in *Proc. SPIE 45th Annu. Meeting*, San Diego, CA, Aug. 2000.

[11] A. Pizurica, W. Philips, I. Lemahieu, and M. Acheroy, "Image de-noising in the wavelet domain using prior spatial constraints," in *Proc. 7th Int. Conf. Image Processing Applications* Manchester, U.K., July 1999, pp. 216–219.

[12] S. G. Mallat and S. Zhong, "Characterization of signals from multiscale edges," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 14, pp. 710–732, July 1992.

[13] S. G. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 11, pp. 674–693, July 1989.

[14] J. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-8, pp. 679–698, 1986.

[15] D. L. Donoho, "Nonlinear wavelet methods for recovery of signals, densities and spectra from indirect and noisy data," in *Proc. Symp. Applied Mathematics*, I. Daubechies, Ed. Providence, RI, 1993.

[16] ——, "Wavelet shrinkage and w.v.d.: A 10-minute tour," *Progr. Wavelet Anal. Applicat.*, 1993.

[17] H. J. Larson and B. O. Shubert, *Probabilistic Models in Engineering Sciences*. New York: Wiley, 1979, vol. I.
[18] M. K. Mihçak, I. Kozintsev, K. Ramchandram, and P. Moulin, "Low-complexity image denoising based on statistical modeling of wavelet coefficients," *IEEE Signal Processing Lett.*, vol. 6, pp. 300–303, Dec. 1999.
[19] H. A. Chipman, E. D. Kolaczyk, and R. E. McCulloch, "Adaptive Bayesian wavelet shrinkage," *J. Amer. Statist. Assoc.*, vol. 92, no. 440, pp. 1413–1421, Dec. 1997.
[20] J. S. Lee, "Digital image enhancement and noise filtering by use of local statistics," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-2, pp. 165–168, Mar. 1980.

Jacob Scharcanski received the B.Eng. degree in electrical engineering in 1981 and the M.Sc. degree in computer science in 1984, both from the Federal University of Rio Grande do Sul, Porto Alegre, RS, Brazil. He received the Ph.D. degree in systems design engineering from the University of Waterloo, Waterloo, ON, Canada, in 1993.

His main areas of interest are image processing and analysis, pattern recognition, and visual information retrieval. He has lectured at the University of Toronto, the University of Guelph, the University of East Anglia, and the University of Manchester, as well as in several Brazilian Universities. He has authored and coauthored more than 60 papers in journals and conferences. He also held research and development positions in the Brazilian and North American Industry. Currently, he is an Associate Professor with the Institute of Informatics, Federal University of Rio Grande do Sul.

Cláudio R. Jung received the B.S. and M.S. degrees in applied mathematics, and the Ph.D. degree in computer science from the Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brazil, in 1993, 1995, and 2002, respectively.

He is an Assistant Professor of mathematics at the Universidade do Vale do Rio dos Sinos, Brazil. His research interests are image filtering, edge detection and enhancement using wavelets, stochastic texture analysis, and image segmentation.

Robin T. Clarke received the M.A. degree in mathematics from the University of Oxford, Oxford, U.K., and the Diploma degree in statistics from the University of Cambridge, Cambridge, U.K. He received the D.Sc. degree from the University of Oxford for his work in hydrology and water resources, a field in which he has since spent much of his career.

He has held appointments (1970–1983) at the Institute of Hydrology of the U.K. Natural Environment Research Council and, from 1983 to 1988, he was Director of the U.K. Freshwater Biological Association. Since 1988, he has held visiting appointments at the Instituto de Pesquisas Hidrulicas, Porto Alegre, Brazil; the University of Plymouth, Plymouth, U.K.; NASA; and the IBM T. J. Watson Research Laboratory. He is the author of three books and several papers on the application of quantitative methods.