

Noise Reduction for Magnetic Resonance Images via Adaptive Multiscale Products Thresholding

Paul Bao* and Lei Zhang

Abstract—Edge-preserving denoising is of great interest in medical image processing. This paper presents a wavelet-based multiscale products thresholding scheme for noise suppression of magnetic resonance images. A Canny edge detector-like dyadic wavelet transform is employed. This results in the significant features in images evolving with high magnitude across wavelet scales, while noise decays rapidly. To exploit the wavelet interscale dependencies we multiply the adjacent wavelet subbands to enhance edge structures while weakening noise. In the multiscale products, edges can be effectively distinguished from noise. Thereafter, an adaptive threshold is calculated and imposed on the products, instead of on the wavelet coefficients, to identify important features. Experiments show that the proposed scheme better suppresses noise and preserves edges than other wavelet-thresholding denoising methods.

Index Terms—Denoising, magnetic resonance image, multiscale products, thresholding, wavelet transform.

I. INTRODUCTION

MAGNETIC resonance imaging (MRI) is a powerful diagnostic technique. However, the incorporated noise during image acquisition degrades the human interpretation, or computer-aided analysis of the images. Time averaging of image sequences aimed at improving the signal-to-noise ratio (SNR) would result in additional acquisition time and reduce the temporal resolution. Therefore, denoising should be performed to improve the image quality for more accurate diagnosis.

Many wavelet transform- [1]–[5] based noise reduction schemes have appeared in literature [18]–[23] during the last two decades, most of which consider the incorporated noise as additive Gaussian white. MRI magnitude images are usually modeled by a *Rician* distribution [9], [10] and the so-called *Rician* noise (the error between the underlying image intensities and the measurement data) is locally signal dependent. The *Rician* noise distribution is well approximated by a Gaussian in bright (high SNR) regions while a Rayleigh distribution is more appropriate in dark (low SNR) regions. A wavelet-based Wiener-filter-like denoising method accounting for the *Rician*

noise was proposed by Nowak [8]. He squared the magnitude MRI image and modeled the square of the *Rician* random variable by a scaled noncentral chi-square distribution. Although the noise in magnitude MRI images are *Rician*, the additive Gaussian white noise assumption holds for each component of the complex MRI data in k -space. Denoising can be applied to the real and imaginary channels, respectively, rather than to the magnitude images. This technique has proved to be effective [11]–[13]. In view of this, the additive Gaussian white noise model is adopted in this paper.

Among the wavelet-based noise reduction techniques, nonlinear thresholding is simple yet very effective. In his innovative work [19], Donoho showed that the *Universal* threshold $t = \sigma\sqrt{2\log N}$ is asymptotically optimal in the *minimax* sense, where σ is the standard deviation of additive white noise and N is the sample length. However, it is well known that the *Universal* threshold over-smoothes images. Donoho improved his work [20] using the *SURE* threshold. It is subband adaptive and is derived by minimizing Stein's unbiased risk estimator. Recently, by modeling the wavelet coefficients within each subband as *i.i.d* random variables with generalized Gaussian distribution (GGD), Chang *et al.* [23] proposed the *BayesShrink* scheme. The *BayesShrink* threshold is also subband dependent and yields better results than the *SURE* threshold. The thresholds mentioned above are based on orthogonal wavelets and are *soft*, implying that the input w is shrunk to zero by an amount of threshold t . In [22], Pan *et al.* presented a *hard* threshold with a nonorthogonal wavelet expansion. Denoting the standard deviation of noise at the j th wavelet scale by σ_j , Pan *et al.* imposed $t_j = c\sigma_j$, where σ_j is the standard deviation of noise at the j th scale and $c \in [3, 4]$ is a constant, to identify significant structures. The word *hard* implies that the input w is preserved if it is greater than the threshold; otherwise it is set to zero. The factor c is a constant and Pan *et al.* set it around 3 since the values of Gaussian distributed noise are, in high probability, within three times its standard deviation. The *soft BayesShrink* and the *hard thresholding* of Pan are used for comparison in the sequel.

There exist dependencies between wavelet coefficients. In [7], Crouse *et al.* used the hidden Markov tree (HMT) models to characterize the joint statistics of wavelet coefficients across scales. In the noise reduction technique of Pizurica [16], the interscale correlation information is exploited to classify the wavelet coefficients. The preliminary classification is then used to estimate the distribution of a coefficient to decide if it is a feature. If a coefficient at a coarser scale has small magnitude, its descendant coefficients at finer scales are likely to be small. Shapiro exploited this property to develop the well-known embedded zerotree wavelet coder [6]. Conversely, if a wavelet

Manuscript received February 25, 2003; revised March 17, 2003. This work was supported in part by the Research Grant Council of the Hong Kong Special Administrative Region under Grant CUHK5982/00E. The Associate Editor responsible for coordinating the review of this paper and recommending its publication was M. Unser. *Asterisk indicates corresponding author.*

*P. Bao is with the Department of Information Engineering, The Chinese University of Hong Kong, Shatin, New Territory, Hong Kong (e-mail: paulbao@ie.cuhk.edu.hk).

L. Zhang is with the Department of Electrical & Computer Engineering, McMaster University, Hamilton, ON L8S 4L8 Canada.

Digital Object Identifier 10.1109/TMI.2003.816958

coefficient produced by a true signal is of large magnitude at a finer scale, its parents at coarser scales are likely to be large as well. However, for those coefficients caused by noise, the magnitudes will decay rapidly along the scales. With this observation, Xu *et al.* [18] multiplied the adjacent wavelet scales to sharpen the important structures while weakening noise. They developed a spatially selective filtering technique by iteratively selecting edge pixels in the multiscale products. Sadler and Swami [24] analyzed the multiscale products and applied them to step detection and estimation. Both of Xu's and Sadler's works are implemented with a dyadic wavelet constructed by Mallat and Zhong [4]. The so-called MZ wavelet is a compactly supported quadratic spline function that approximates the first derivative of Gaussian. The corresponding dyadic wavelet transform (DWT) is equivalent to the Canny edge detection [17] and characterizes the instantaneous features in a signal well. The MZ wavelet is also employed in this paper.

Wavelet thresholding is simple and efficient but takes no advantage of the dependency information between wavelet scales. In this paper, we present a multiscale thresholding scheme to incorporate the merits of interscale dependencies into the thresholding technique for denoising. Two adjacent wavelet subbands are multiplied to amplify the significant features and dilute noise. In contrast to other schemes, we apply thresholding to the multiscale products instead of the wavelet coefficients. As we will show, the presented multiscale products thresholding can distinguish edge structures from noise more effectively. The variance of noise needs to be estimated to implement the denoising scheme. A new noise level estimator is also proposed in this paper.

This paper is organized as follows. Section II discusses the wavelet multiscale products. Section III describes the thresholding scheme. An image adaptive threshold imposed on the multiscale products is calculated to identify the significant structures. Experiments are given in Section IV in comparison with some wavelet thresholding schemes. The paper is concluded in Section V.

II. WAVELET MULTISCALE PRODUCTS

A. Dyadic Wavelet Transform as a Multiscale Edge Detector

A wavelet transform represents a signal f as a linear combination of elementary atoms that appear at different resolutions. It is computed by convoluting the input signal with dilated wavelet filters recursively. More details about the theory of wavelets and their applications in signal processing can be found in Daubechies [1], Meyer [2], Mallat [3], [4], and Vetterli [5].

We denote by $\xi_s(x)$ the dilation of a function $\xi(x)$ by a scale factor s

$$\xi_s(x) = \frac{\xi\left(\frac{x}{s}\right)}{s}. \quad (2.1)$$

Suppose function $\psi(x)$ satisfies the requirements to be a wavelet. The continuous wavelet transform of any measurable

and square-integrable function $f(x)$, $f \in L^2(R)$, at scale s and position x is defined as

$$W_s f(x) = f * \psi_s(x) \quad (2.2)$$

where the symbol $*$ denotes the convolution operation.

The wavelet transform can be designed as a multiscale edge detector to enhance the signal's instantaneous features [4]. Suppose that $\theta(x)$ is a differentiable smooth function whose integral is equal to 1 and that it converges to 0 at infinity. Let's define $\psi(x)$ as the first-order derivative of $\theta(x)$

$$\psi(x) = \frac{d\theta(x)}{dx}. \quad (2.3)$$

Then, $W_s f(x)$ can be written as

$$W_s f(x) = f * \left(s \frac{d\theta_s}{dx} \right) (x) = s \frac{d}{dx} (f * \theta_s)(x). \quad (2.4)$$

It can be seen that the wavelet transform $W_s f(x)$ is the first derivative of $f(x)$ smoothed by $\theta_s(x)$. In particular, when $\theta(x)$ is a Gaussian function, the local extrema determination in $W_s f(x)$ is equivalent to the well-known Canny edge detection [17].

The Canny edge detector-like wavelet transform can be extended to two-dimensional (2-D) images. Suppose $\theta(x, y)$ is a 2-D differentiable smooth function whose integral is equal to 1 and converges to 0 at infinity. For example $\theta(x, y)$ could be the tensor product of one-dimensional (1-D) smooth functions: $\theta(x, y) = \theta(x) \cdot \theta(y)$. We define the two wavelets $\psi^x(x, y)$ and $\psi^y(x, y)$ at horizontal and vertical directions as

$$\psi^x(x, y) = \frac{\partial \theta(x, y)}{\partial x}, \quad \psi^y(x, y) = \frac{\partial \theta(x, y)}{\partial y}. \quad (2.5)$$

The dilation of any 2-D function $\xi(x, y)$ by scale s can be, therefore, denoted by

$$\xi_s(x, y) = s^{-2} \xi\left(\frac{x}{s}, \frac{y}{s}\right). \quad (2.6)$$

Suppose $f(x, y)$ is a 2-D measurable and square-integrable function such that $f \in L^2(R^2)$. The wavelet transform of $f(x, y)$ at scale s and position (x, y) has two components

$$\begin{aligned} W_s^x f(x, y) &= f * \psi_s^x(x, y) \\ \text{and } W_s^y f(x, y) &= f * \psi_s^y(x, y). \end{aligned} \quad (2.7)$$

Similarly to (2.4) these two components can be rewritten as

$$\begin{aligned} W_s^x f(x, y) &= s \frac{\partial}{\partial x} (f * \theta_s)(x, y) \\ \text{and } W_s^y f(x, y) &= s \frac{\partial}{\partial y} (f * \theta_s)(x, y). \end{aligned} \quad (2.8)$$

In the case when $\theta(x, y)$ is a Gaussian function, detecting the local extrema form $\begin{pmatrix} W_s^x f(x, y) \\ W_s^y f(x, y) \end{pmatrix}$ is equivalent to the Canny edge detection.

For the purpose of fast numerical implementation, we restrict the scale s to vary along the dyadic sequence $(2^j)_{j \in \mathbb{Z}}$. For sim-

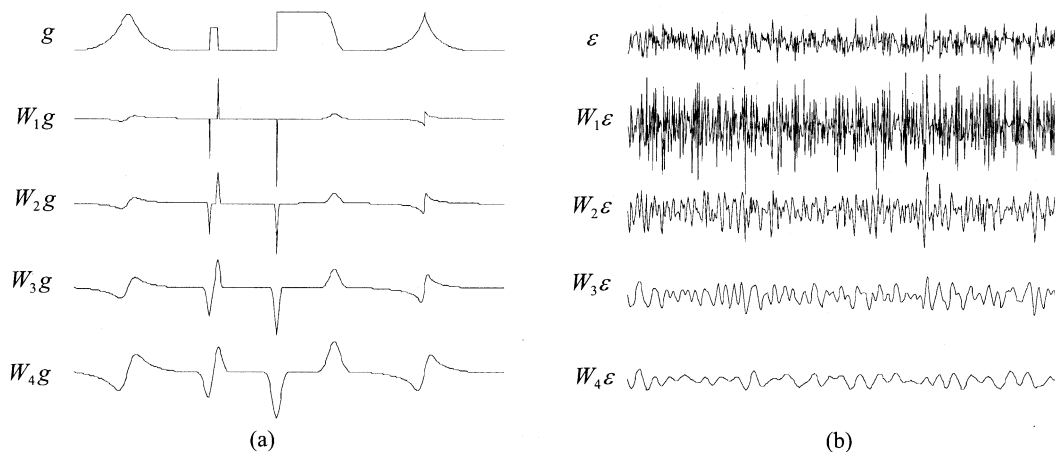


Fig. 1. (a) DWT of a test signal g at the first four scales. (b) The DWT of a sequence of *Gaussian* white noise ε at the first four scales.

plicity, we denote by $\xi_j(x)$ (no confusion with $\xi_s(x)$ in (2.1) the dilation of function $\xi(x)$ by 2^j , then

$$\xi_j(x) = \frac{\xi\left(\frac{x}{2^j}\right)}{2^j}. \quad (2.9)$$

The DWT of $f(x)$ at dyadic scale 2^j and position x is

$$W_j f(x) = f * \psi_j(x). \quad (2.10)$$

The function $f(x)$ can be recovered from its DWT by

$$f(x) = \sum_{j=-\infty}^{\infty} W_j f * \chi_j(x) \quad (2.11)$$

where $\chi(x)$ is any reconstructing wavelet whose Fourier transform (FT) satisfies [4]

$$\sum_{j=-\infty}^{\infty} \hat{\psi}(2^j \omega) \hat{\chi}(2^j \omega) = 1. \quad (2.12)$$

The wavelet used in this paper is the MZ wavelet constructed by Mallat and Zhong [4]. The associated smooth function $\theta(x)$ is a cubic spline, which closely approximates a Gaussian function. The wavelet $\psi(x)$ is a quadratic spline that approximates the first derivative of Gaussian. Thus, the DWT behaves like a Canny edge detector. Appendix A shows the functions $\theta(x)$ and $\psi(x)$, and illustrates the discrete decomposition algorithms of the 1-D and 2-D DWT. Details about the derivation of the MZ wavelet can be found in [4].

B. Multiscale Products

Signals and noise behave very differently in the wavelet transform domain. The evolution of singularities and noise across wavelet scales were analyzed by Mallat *et al.* [3], [4] using the mathematical concept of the Lipschitz regularity. Singularities are more regular than noise and have higher Lipschitz regularities. For example, the Lipschitz regularity of a step edge is 0. If a structure is smoother than the step, it will have positive Lipschitz regularity. Otherwise it can be considered having negative Lipschitz regularity. The Lipschitz regularity of the Dirac function is equal to -1 . White noise is almost singular everywhere and has a uniform Lipschitz regularity that is equal to $-1/2$.

Meyer [2] presented a theorem to relate the evolution of the wavelet transform magnitude with the signal's Lipschitz regularity. A function $f(x)$ is uniformly Lipschitz α ($0 < \alpha < 1$) over interval $[a, b]$ if and only if there exists a constant $K > 0$ such that for all $x \in [a, b]$, the wavelet transform satisfies

$$|W_j f(x)| \leq K(2^j)^\alpha. \quad (2.13)$$

The above equation implies that the wavelet transform magnitudes increase for positive α with increasing scales. Contrarily, wavelet transform magnitudes decrease for negative Lipschitz regularities with increasing scales. In Fig. 1, the DWT at the first four scales of a test signal g , and a sequence of Gaussian white noise ε , are illustrated. Notice that the signal singularities evolve across scales with observable peaks while noise decays rapidly along scales. As illustrated in Fig. 1(b), Mallat and Hwang [3] observed that, for Gaussian white noise, the average number of local maxima at scale 2^{j+1} is half of that at scale 2^j .

With the observation of Fig. 1, we can imagine that multiplying the DWT at adjacent scales would amplify edge structures and dilute noise. This favorite property has been exploited by Xu *et al.* [18] and Sadler [24] in noise reduction and step detection. In this paper, we define the multiscale products of $W_j f$ as

$$P_j f(x) = \prod_{i=-k_1}^{k_2} W_{j+i} f(x) \quad (2.14)$$

where k_1 and k_2 are nonnegative integers.

The support of an isolated edge will increase by a factor of two across scale and the neighboring edges will interfere with each other at coarse scales (Fig. 1). So in practice it is sufficient to implement the multiplication at two adjacent scales. Let $k_1 = 0$ and $k_2 = 1$, then we calculate the DWT scale products as

$$P_j f(x) = W_j f(x) \cdot W_{j+1} f(x). \quad (2.15)$$

Similarly for 2-D images, the multiscale products have two components

$$\begin{aligned} P_j^x f(x, y) &= W_j^x f(x, y) \cdot W_{j+1}^x f(x, y) \\ \text{and } P_j^y f(x, y) &= W_j^y f(x, y) \cdot W_{j+1}^y f(x, y). \end{aligned} \quad (2.16)$$

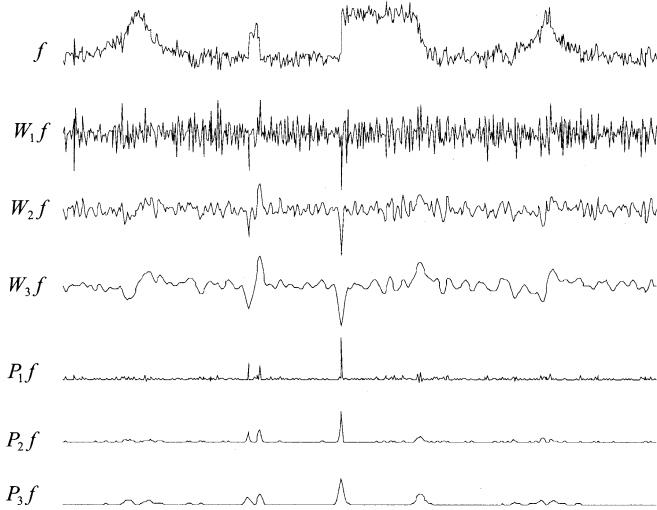


Fig. 2. The DWT and multiscale products of a noisy test signal at the first three scales.

In Fig. 2, the DWT and multiscale products of a noisy test signal $f = g + \varepsilon$ are illustrated. Though the wavelet transform coefficients of the original signal g are immersed into noise at fine scales, they are enhanced in the scale products $P_j f$. The significant features of g are more distinguishable in $P_j f$ than in $W_j f$.

III. ADAPTIVE MULTISCALE PRODUCTS THRESHOLDING

A. The Thresholding Scheme

Wavelet-based thresholding techniques have proved to be effective in denoising [18]–[23]. Nonsignificant wavelet coefficients below a preset threshold value are discarded as noise and the image is reconstructed from the remaining significant coefficients. Compared with the linear denoising methods that blur images as well as smoothing noise, the nonlinear wavelet thresholding schemes preserve image singularities better.

In general, thresholds are classified into *soft* and *hard*. The *soft* thresholds shrink the input wavelet coefficient w to zero by an amount t , i.e., $\eta_t(w) = \text{sgn}(w) \cdot \max(0, |w| - t)$. Contrarily, the *hard* thresholds preserve the input coefficient if it is greater than the threshold, i.e., $\eta_t(w) = w \cdot \mathbf{1}\{|w| > t\}$. Naturally, the determination of the threshold value is extremely critical to the threshold-based algorithms. We denote by $f = g + \varepsilon$ the measurements of image g corrupted by Gaussian white noise $\varepsilon \sim N(0, \sigma^2)$. Donoho *et al.* [19] presented the *Universal threshold* $t = \sigma\sqrt{2\log N}$ in his well-known *Wavelet Shrinkage* scheme. Chang *et al.* [23] presented the *BayesShrink* threshold $t = \sigma^2/\sigma_{W_j g}$, where $\sigma_{W_j g}$ is the image standard deviation at the j th wavelet scale. The above three thresholds are *soft* and are derived from orthogonal wavelet bases. A *hard* threshold that could be applied to nonorthogonal wavelet transforms was proposed by Pan *et al.* [22].

All the above wavelet thresholding schemes impose the threshold directly on wavelet coefficients. They do not exploit the dependencies that exist between adjacent wavelet scales. From Fig. 2, it can be noticed that, at finer scales, if the threshold t applied to $W_j f$ is relatively sizeable, some edge

structures may be suppressed as noise. Otherwise, if t is relatively small, many noisy pixels would be undesirably preserved. However, in the multiscale products $P_j f$ it can be seen that the significant structures are strengthened while the noise is weakened. $P_j f$ results in a more effective discrimination between edges and noise than $W_j f$. With such observations Xu *et al.* [18] and Sadler *et al.* [24] have exploited multiscale products in denoising and step estimation.

In this paper, we propose a new de-noising scheme, the *adaptive multiscale products thresholding*, to merge the merits of the thresholding technique and wavelet interscale dependencies. A significant wavelet coefficient $\hat{W}_j^d f(x, y)$, where $d = x, y$ indicates x or y dimension, is identified if its corresponding multiscale products value $P_j^d f(x, y)$ is greater than an adaptive threshold $t_p^d(j)$. The algorithm is summarized as follows.

1) Compute the DWT of input image f up to J scales.

2) Calculate the multiscale products $P_j^d f$ and preset the thresholds $t_p^d(j)$. Then threshold the wavelet coefficients by

$$\hat{W}_j^d f(x, y) = \begin{cases} W_j^d f(x, y) & P_j^d f(x, y) \geq t_p^d(j) \\ 0 & P_j^d f(x, y) < t_p^d(j) \end{cases}, \quad j = 1, \dots, J; d = x, y. \quad (3.1)$$

3) Recover the image from the thresholded wavelet coefficients $\hat{W}_j^x f(x, y)$ and $\hat{W}_j^y f(x, y)$.

B. Determination of the Threshold

Since a wavelet transform is a linear transform, the DWT of a noisy image $f = g + \varepsilon$ can be written as

$$W_j^d f = W_j^d g + W_j^d \varepsilon \quad (3.2)$$

where $W_j^d g$ is the DWT of original image g and $W_j^d \varepsilon$ is the DWT of additive noise ε . For convenience, we denote

$$Z_j^d = P_j^d f = W_j^d f \cdot W_{j+1}^d f. \quad (3.3)$$

Due to the high dependencies existing between $W_j^d f$ and $W_{j+1}^d f$, the histograms of Z_j^d will have a heavy positive tail (See Fig. 4). A proper threshold $t_p^d(j)$ can be determined and imposed on Z_j^d to eliminate the highly noise corrupted pixels and identify the significant image structures.

Suppose that the input image is Gaussian white noise ε and it is an ergodic stationary process. For the convenience of expression, we denote the DWT of ε by

$$U_j^*(x, y) = W_j^* \varepsilon(x, y) = \varepsilon * \psi_j^*(x, y), \quad * = x, y. \quad (3.4)$$

U_j^* is a Gaussian colored noise process and its standard deviation is

$$\sigma_j = \|\psi_j\| \sigma \quad (3.5)$$

where norm $\|\psi_j\| = \sqrt{\iint \psi_j^2(x, y) dx dy}$. We do not use the superscript “ d ” in (3.5) because the norm values $\|\psi_j^x\|$ and $\|\psi_j^y\|$ are the same.

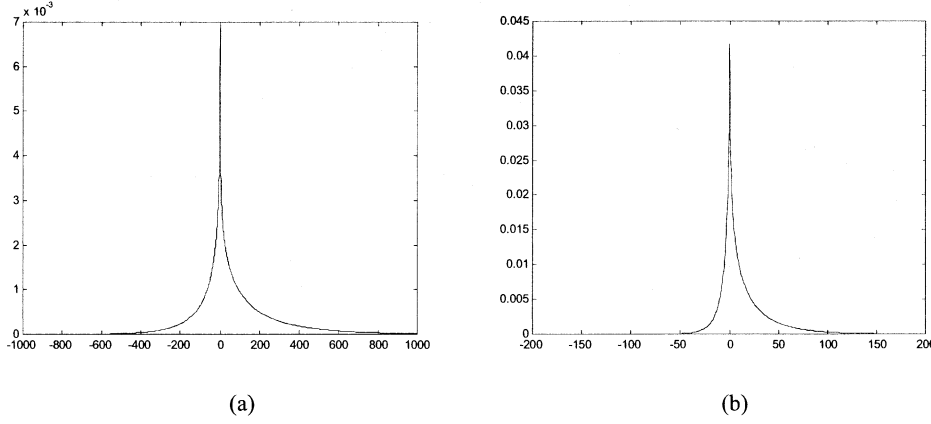


Fig. 3. The theoretical pdf of the multiscale products V_j for *Gaussian* white noise ($\sigma = 10$). (a) At the first scale $j = 1$. (b) At the second scale $j = 2$.

U_j^d and U_{j+1}^d are jointly Gaussian distributed with probability density function (pdf) [26]

$$p(u_j, u_{j+1}) = \frac{1}{2\pi\sigma_j\sigma_{j+1}\sqrt{1-\rho_{j+1,j}^2}} \cdot e^{-1/2(1-\rho_{j+1,j}^2)[u_j^2/\sigma_j^2 - (2\rho_{j+1,j}u_ju_{j+1}/\sigma_j\sigma_{j+1}) + u_{j+1}^2/\sigma_{j+1}^2]} \quad (3.6)$$

where the correlation coefficient $\rho_{j+1,j}$ of U_j^d and U_{j+1}^d is

$$\rho_{j+1,j} = \frac{\int \int \psi_j(x, y) \cdot \psi_{j+1}(x, y) dx dy}{\sqrt{\int \int \psi_j^2(x, y) dx dy \cdot \int \int \psi_{j+1}^2(x, y) dx dy}}. \quad (3.7)$$

The values of σ_j and $\rho_{j+1,j}$ in the discrete implementation are listed in Table I by setting $\sigma = 1$.

We denote the scale products of U_j^d and U_{j+1}^d by

$$V_j^d = U_j^d \cdot U_{j+1}^d, \quad (3.8)$$

Then the pdf of V_j^d will have the following form [25, pp. 42]:

$$p(v_j) = \frac{1}{\pi\Gamma(\frac{1}{2})\sigma_j\sigma_{j+1}\sqrt{1-\rho_{j+1,j}^2}} \cdot e^{(\rho_{j+1,j}v_j/(1-\rho_{j+1,j}^2)\sigma_j\sigma_{j+1})} K_0\left(\frac{|v_j|}{(1-\rho_{j+1,j}^2)\sigma_j\sigma_{j+1}}\right) \quad (3.9)$$

where $\Gamma(t) = \int_0^\infty e^{-u}u^{t-1}du$ is the Gamma function and K_0 is the modified Bessel function of the second kind with order zero. When $\rho_{j+1,j}$ is positive, $p(v_j)$ is right skewed. In Fig. 3, the theoretical pdf's $p(v_1)$ and $p(v_2)$ are plotted by setting $\sigma = 10$. Notice that $p(v_2)$ is more positively tailed than $p(v_1)$ because $\rho_{3,2}$ is higher than $\rho_{2,1}$.

In applications, the wavelet coefficient obtained is $W_j^d f$, which is the sum of noiseless coefficient $W_j^d g$ and noise $W_j^d \varepsilon$. Since white noise is singular almost everywhere, at fine scales $W_j^d \varepsilon$ will be predominant in $W_j^d f$ except for some significant features to be preserved (see Fig. 2). In Fig. 4(a), the histograms of Z_j^x at the first three scales are plotted for the noiseless image *Lena* (i.e., $Z_j^x = W_j^x g \cdot W_{j+1}^x g$). Fig. 4(b) shows the histograms of Z_j^x when the input is Gaussian white noise with zero mean and standard deviation $\sigma = 30$ (i.e.,

$Z_j^x = W_j^x \varepsilon \cdot W_{j+1}^x \varepsilon$). In Fig. 4(c), the histograms of Z_j^x , where the input is the noisy *Lena* ($\sigma = 30$, SNR = 12.93 dB), are shown (i.e., $Z_j^x = W_j^x f \cdot W_{j+1}^x f$). It is noticeable that at scales 2^1 and 2^2 , the corresponding histograms in Fig. 4(b) and (c) are very similar. This is because the energy of noise in these subbands is relatively high. At coarse scales the energy of the image increases, but that of noise decreases rapidly. The histograms of $Z_j^x = W_j^x f \cdot W_{j+1}^x f$ will be close to those of $Z_j^x = W_j^x g \cdot W_{j+1}^x g$ step by step.

The standard deviation of V_j^d is [26]

$$\kappa_j = \sqrt{E[v_j^2]} = \sqrt{E[u_j^2 u_{j+1}^2]} = \sqrt{1 + 2\rho_{j+1,j}^2} \cdot \sigma_j \sigma_{j+1}. \quad (3.10)$$

In Table II, we compute the values of probability

$$\Pr_j(c) = P\{v_j \leq c \cdot \kappa_j\} \quad (3.11)$$

where constant c varies from 1 to 5 by step length 1. Notice that when $c \geq 5$, the probability $\Pr_j(c) \rightarrow 1$, implying that $5\kappa_j$ will suppress most of the data in V_j .

We denote

$$\mu_f^d(j) = E[Z_j^d], \mu_\varepsilon^d(j) = E[V_j^d]$$

and

$$\mu_g^d(j) = E[W_j^d g \cdot W_{j+1}^d g]. \quad (3.12)$$

Since the noise ε is independent of the noiseless image g , it can be derived that

$$\mu_g^d(j) = \mu_f^d(j) - \mu_\varepsilon^d(j) \quad (3.13)$$

and

$$\mu_\varepsilon^d(j) = \rho_{j+1,j} \sigma_j \sigma_{j+1}. \quad (3.14)$$

The ratio $\mu_\varepsilon^d(j)/\mu_g^d(j)$ is a measurement for the intensity of noise against signal in the multiscale products Z_j^d . This ratio can be used to adjust the threshold $t_p^d(j)$ imposed on Z_j^d . We set the multiscale products threshold as

$$t_p^*(j) = 5\kappa_j \left(1 + \frac{\mu_\varepsilon^*(j)}{\mu_g^*(j)}\right). \quad (3.15)$$

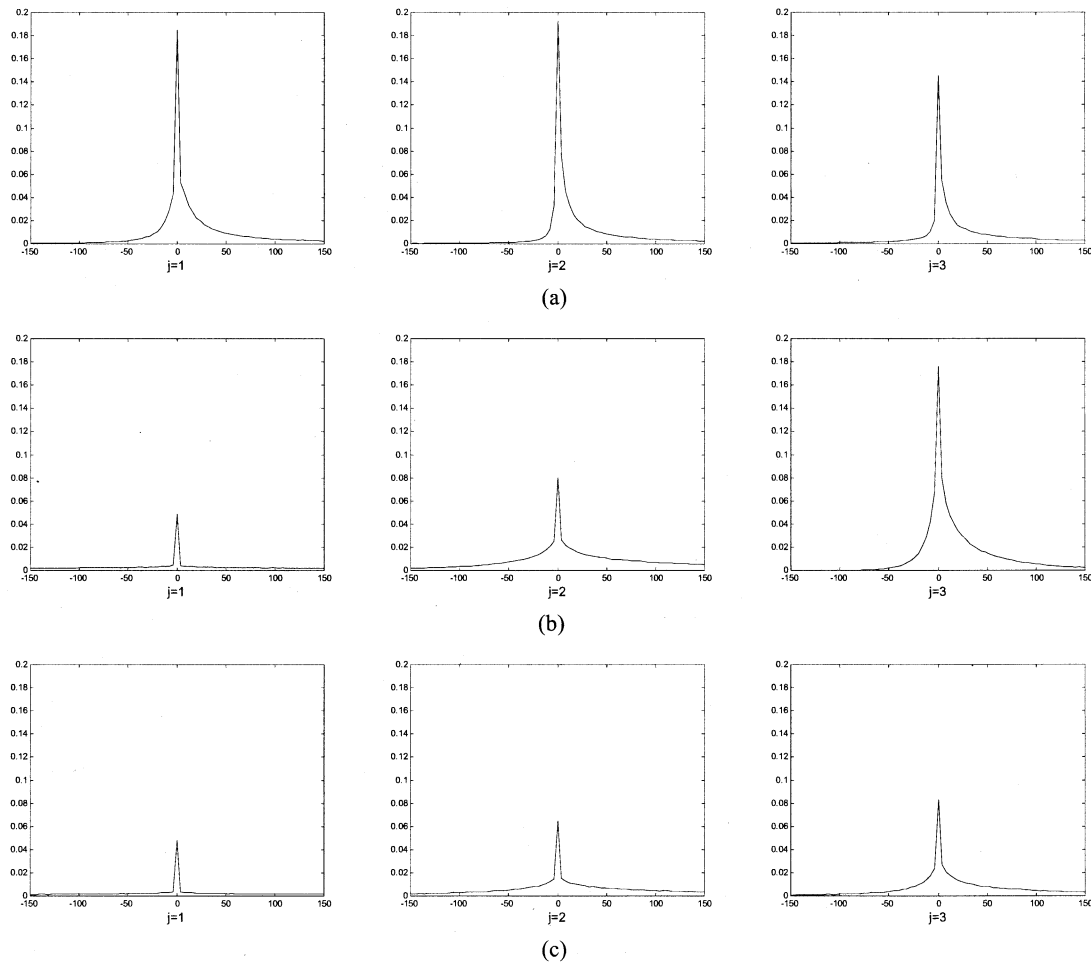


Fig. 4. The Histograms of the multiscale products $P_j^x f$ at the first three scales when f is (a) Noiseless image *Lena*; (b) Gaussian white noise; (c) Noisy image *Lena*.

The adaptive threshold $t_p^d(j)$ is intuitive and effective. When noise is much stronger compared with the image (i.e., at fine scales), the ratio $\mu_\varepsilon^d(j)/\mu_g^d(j)$ is high. Therefore, the threshold $t_p^d(j)$ becomes sufficiently large to suppress the overwhelming noise. When the image is dominative (i.e., at coarse scales or when additive noise is low), the ratio $\mu_\varepsilon^d(j)/\mu_g^d(j)$ is small and the threshold $t_p^d(j)$ is at an appropriate level to preserve the image instantaneous features while removing noise. In Table III, we give some values of $\mu_\varepsilon^d(j)/\mu_g^d(j)$ for image *Lena* with noise levels $\sigma = 10, 20, 30$, and 40 , respectively.

C. Noise Level Estimation

The standard deviation of additive Gaussian white noise, σ , should be estimated to implement the denoising scheme. A popular noise level estimator has been proposed by Donoho [19]. The median absolute value (MAV) of the wavelet coefficients at the finest scale is first calculated and the standard deviation of noise is then estimated as $\text{MAV}/0.6745$. The MAV estimator is inaccurate for those images containing massive fine structures.

We propose a new noise level estimation method here. We compute the orthogonal wavelet transform (OWT) of the noisy image at the finest scale and denote by W the wavelet coefficients in the diagonal direction. Because OWT is a unitary trans-

TABLE I
NOISE STANDARD DEVIATION AND CORRELATION COEFFICIENT VALUES OF THE MZ WAVELET IN DISCRETE IMPLEMENTATION AT SCALE 2^j . INPUT NOISE IS ASSUMED TO BE UNIT GAUSSIAN WHITE

j	1	2	3	4
σ_j	2.8284	0.7395	0.3173	0.1531
$\rho_{j+1,j}$	0.3586	0.5504	0.5957	0.6063

TABLE II
THE VALUES OF PROBABILITY $\text{Pr}_j(c) = P\{v_j \leq c \cdot \kappa_j\}$ FOR $j = 1$ AND $j = 2$

c	1	2	3	4	5
$\text{Pr}_1(c)$	0.8445	0.9456	0.9799	0.9927	0.9976
$\text{Pr}_2(c)$	0.8291	0.9396	0.9774	0.9915	0.9971

TABLE III
THE RATIOS OF $\mu_\varepsilon^x(j)/\mu_g^x(j)$ AT THE FIRST TWO SCALES FOR IMAGE *LENA* ON DIFFERENT NOISE LEVELS σ

σ	10	20	30	40
$\mu_\varepsilon^x(1)/\mu_g^x(1)$	0.27	1.18	3.20	7.33
$\mu_\varepsilon^x(2)/\mu_g^x(2)$	0.02	0.10	0.22	0.40

TABLE IV
NOISE LEVEL ESTIMATION RESULTS. GAUSSIAN WHITE NOISE WITH STANDARD DEVIATION σ IS ADDED TO FIVE TEST IMAGES. $\hat{\sigma}_d$ IS THE ESTIMATION BY DONOHO'S MEDIAN METHOD AND $\hat{\sigma}$ IS THE ESTIMATION BY OUR SCHEME

σ		5	10	15	20	25	30	35	40
<i>Lena</i>	$\hat{\sigma}_d$	6.15	11.02	15.87	20.82	25.68	30.56	35.54	40.52
	$\hat{\sigma}$	3.20	8.37	14.30	19.66	24.13	30.21	35.31	40.56
<i>Sailboat</i>	$\hat{\sigma}_d$	7.84	12.16	16.78	21.47	26.20	31.13	35.94	40.86
	$\hat{\sigma}$	4.06	8.78	14.57	19.88	25.14	30.71	35.83	40.80
<i>Goldhill</i>	$\hat{\sigma}_d$	7.49	11.79	16.46	21.18	25.94	30.84	35.69	40.67
	$\hat{\sigma}$	4.22	9.02	14.70	20.13	25.34	30.42	35.21	40.78
<i>Bridge</i>	$\hat{\sigma}_d$	8.46	12.73	17.17	21.84	26.54	31.29	36.19	40.98
	$\hat{\sigma}$	4.89	9.58	15.10	20.51	25.82	30.95	35.48	40.94
<i>Baboon</i>	$\hat{\sigma}_d$	14.58	18.17	22.29	26.65	30.93	35.34	39.90	44.51
	$\hat{\sigma}$	7.05	9.66	13.61	18.66	24.01	29.53	35.25	41.37

form, at each wavelet scale the noise standard deviation is equal to σ . Thus, the variance of W is

$$\sigma_f^2 = E[W^2] = \sigma_g^2 + \sigma^2 \quad (3.16)$$

where σ_g is the standard deviation of the wavelet coefficients of the noiseless image.

Suppose N is a zero-mean Gaussian process with standard deviation σ_N . We divide it into two parts. The first part N_a consists of points that $|N(\cdot)| > \sigma_N$ and the second part N_b consists of points where $|N(\cdot)| \leq \sigma_N$. Let $\sigma_a = \sqrt{E[N_a^2]}$ and $\sigma_b = \sqrt{E[N_b^2]}$, we have

$$\sigma_a^2 = 2 \int_{\sigma_N}^{\infty} \frac{x^2}{\sqrt{2\pi}\sigma_N(1 - 2\text{erf}(1))} e^{-x^2/(2\sigma_N^2)} dx \quad (3.17)$$

$$\sigma_b^2 = \int_0^{\sigma_N} \frac{x^2}{\sqrt{2\pi}\sigma_N \text{erf}(1)} e^{-x^2/(2\sigma_N^2)} dx \quad (3.18)$$

where $\text{erf}(t) = 1/\sqrt{2\pi} \int_0^t e^{-x^2/2} dx$ is the error function. The ratio of σ_a to σ_b is independent of σ_N and we can calculate that $\sigma_a/\sigma_b \approx 2.945$.

Next, we split W into two parts, W_a such that $|W_a(\cdot)| > \sigma_f$ and W_b such that $|W_b(\cdot)| \leq \sigma_n$. Let $\sigma_w^a = \sqrt{E[W_a^2]}$ and $\sigma_w^b = \sqrt{E[W_b^2]}$. Generally, the noise energy is concentrated on W_b and σ_w^b can be considered as an approximation of the noise level σ . We define

$$\hat{\sigma}_g = \sqrt{(\sigma_w^a)^2 - 2.9^2(\sigma_w^b)^2}. \quad (3.19)$$

If W is produced totally by noise, obviously $\hat{\sigma}_g$ will be equal to zero. The more image details involved in W_a , the greater the value of $\hat{\sigma}_g$. In this case, $\hat{\sigma}_g$ can be seen as an approximated estimation of σ_g . We then take

$$r = \frac{\hat{\sigma}_g}{\sigma_w^b} \quad (3.20)$$

to be a measure of SNR at the finest scale. Finally, the noise level can be estimated as

$$\hat{\sigma} = \frac{\sigma_f}{\sqrt{1+r^2}}. \quad (3.21)$$

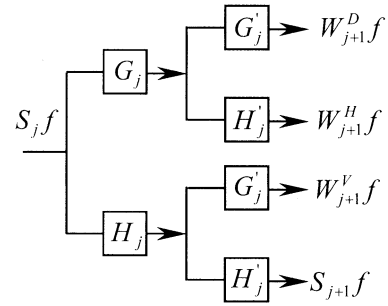


Fig. 5. One stage transform structure of the 2-D overcomplete wavelet expansion (OWE). Filter F_j is the 2^j scale dilation of F_0 (putting $2^j - 1$ zeros between each of the coefficients of F_0) and F'_j is the transpose of F_j . $W_j^H f$, $W_j^V f$ and $W_j^D f$ are the wavelet coefficients at the horizontal, vertical, and diagonal directions.

We denote the MAV noise estimator of Donoho by

$$\hat{\sigma}_d = \frac{\text{Median}(|W|)}{0.6745}. \quad (3.22)$$

The Monte Carlo experimental results using the two estimators are listed in Table IV. The test images employed are *Lena*, *Sailboat*, *Goldhill*, *Bridge*, and *Baboon*. We added the Gaussian white noise with different standard deviation σ to each of them. The Daubechies wavelet [1] with four vanishing moments is used for the OWT. It can be observed that the proposed method generally outperforms Donoho's MAV estimation scheme. It performs especially well for the image *Baboon* that contains massive fine structures.

IV. EXPERIMENTS

In this section, the performances by the proposed scheme on some MRI images are compared with those of the *soft* thresholding scheme *BayesShrink* of Chang *et al.* [23] and the *hard* thresholding scheme of Pan *et al.* [22]. For convenience, we refer the two methods as *STH* and *HTH*, respectively. It is well observed that thresholding with the OWT produces unpleasant Gibbs-like edge artifacts [21]. Thus, we implement the two schemes with the over-complete wavelet expansion (OWE). The one stage transform of OWE is illustrated in Fig. 5. The resultant denoising by thresholding with the OWE can be interpreted as the average of the circularly shifted denoising

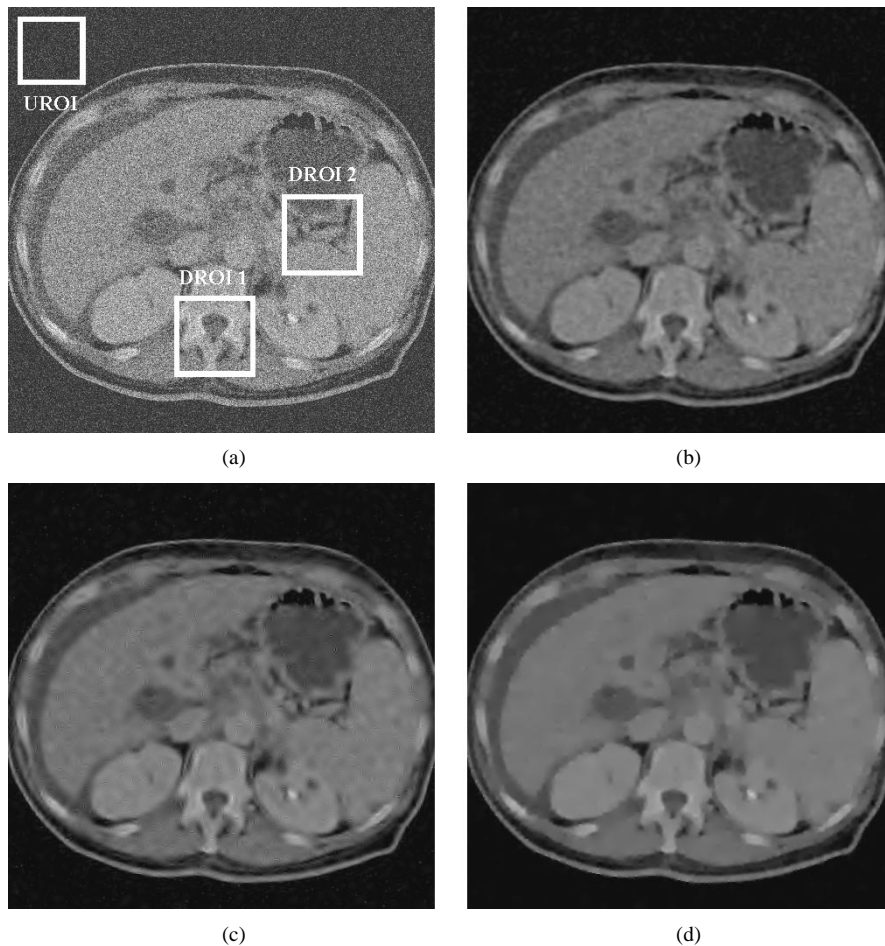


Fig. 6. Experiments on MRI image *Liver*. The DROI and UROI used to compute the MSR and CNR indexes (listed in Table V) are highlighted. (a) The noisy image. (b) Estimated by the *STH*. (c) Estimated by the *HTH*. (d) Estimated by the presented *MPTH*.

outcomes by the OWT. The residual noise is better smoothed and the artifacts are attenuated. The wavelet employed in the *STH* and *HTH* schemes is the compactly supported orthogonal wavelet of Daubechies with four vanishing moments [1]. The constant c appearing in the threshold of the scheme *HTH* is set at 3.1. The proposed scheme is referred as *MPTH*. The MRI images in our experiments are 512×512 in size and the decomposition level is four.

To evaluate the medical image quality, we compute the mean-to-standard-deviation ratio (MSR) [14], [15] in a desired region of interest (DROI)

$$\text{MSR} = \frac{\mu_d}{\sigma_d} \quad (4.1)$$

where μ_d and σ_d are the mean and the standard deviation computed in the DROI. The contrast to noise ratio (CNR) is also an important quality measurement for medical image interpretation. It is defined as

$$\text{CNR} = \frac{|\mu_d - \mu_u|}{\sqrt{0.5(\sigma_d^2 + \sigma_u^2)}} \quad (4.2)$$

where μ_u and σ_u are the mean and the standard deviation computed in an undesired region of interest (UROI) such as a window or background. Both the MSR and CNR measurements are proportional to the medical image quality.

In the k -space, the raw MRI data is denoted by [8]–[10]

$$F(\mu, \nu) = G(\mu, \nu) + \xi(\mu, \nu) \quad (4.3)$$

where $G(\mu, \nu)$ is the underlying signal and $\xi(\mu, \nu)$ is a complex Gaussian white noise. By computing the FT of $F(\mu, \nu)$ in the complex image domain, we have

$$f(x, y) = g(x, y) + \varepsilon(x, y). \quad (4.4)$$

$\varepsilon(x, y)$ is again a complex Gaussian white noise due to the unitarity of the FT. The denoising schemes could be applied to each of the real and imaginary components of $f(x, y)$. For visual inspection, the moduli of the complex data $f(x, y)$ are shown as the magnitude image.

Fig. 6(a) is a noisy MRI image *Liver*. The DROI and UROI used for calculating the MSR and CNR indexes are highlighted. Denoised images by the three schemes are illustrated in Fig. 6(b)–(d), respectively, and the MSR and CNR values are listed in Table V. The presented algorithm *MPTH* achieves the highest quantity measurements. Notice that the denoised image by the *STH* contains a few stains and the result by the *HTH* retains much noise. (If the threshold of the *HTH* is set higher to suppress noise, the estimated image would be over-smoothed.) The *MPTH* preserves edges better and yet effectively removes noise. Zoom-in images of DROI 1 and DROI 2 are illustrated

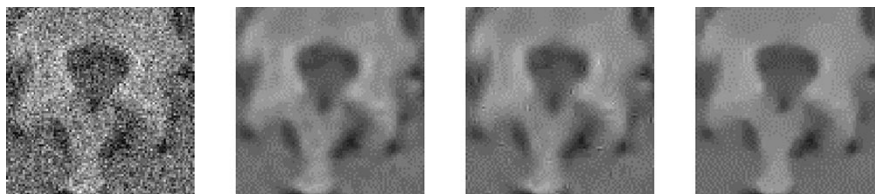


Fig. 7. Zoom in of the DROI 1. (a) The noisy image. (b) Estimated by the *STH*. (c) Estimated by the *HTH*. (d) Estimated by the presented *MPTH*.

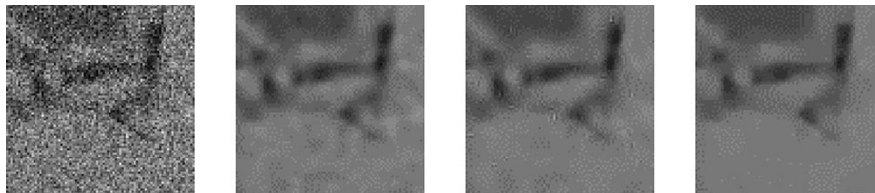


Fig. 8. Zoom in of the DROI 2. (a) The noisy image. (b) Estimated by the *STH*. (c) Estimated by the *HTH*. (d) Estimated by the presented *MPTH*.

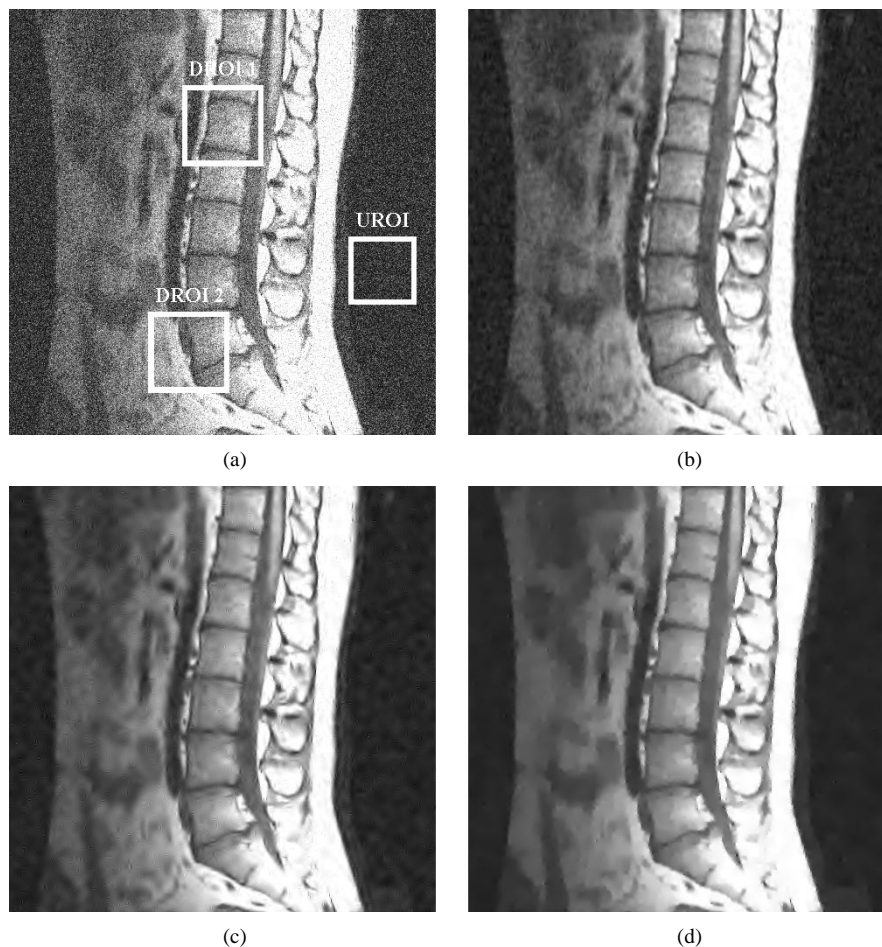


Fig. 9. Experiments on MRI image *Spine*. The DROI and UROI used to compute the MSR and CNR indexes (listed in Table VI) are highlighted. (a) The noisy image. (b) Estimated by the *STH*. (c) Estimated by the *HTH*. (d) Estimated by the presented *MPTH*.

in Figs. 7 and 8. Another experiment on an MRI image *Spine* is illustrated in Figs. 9–11. The MSR and CNR measurements are listed in Table VI. The results showed in Fig. 9(b) and (c) appear to be veiled by the residual noise. Zoom-in images of the two DROI are illustrated in Figs. 10 and 11. Although some *stings* (discontinuities) appeared in Figs. 10(d) and 11(d), they are almost edge points detected from the multiscale products, but discarded by the other two schemes.

V. CONCLUSION

This paper proposes an MRI image denoising scheme using an adaptive wavelet thresholding technique. Unlike many traditional schemes that directly threshold the wavelet coefficients, the proposed scheme multiplies the adjacent wavelet subbands to amplify the significant features and then applies the thresholding to the multiscale products to better differentiate edge

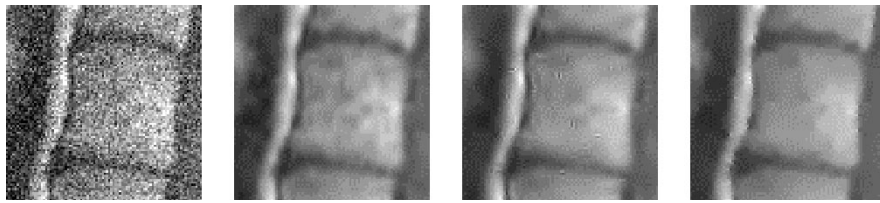


Fig. 10. Zoom in of the DROI 1. (a) The noisy image. (b) Estimated by the *STH*. (c) Estimated by the *HTH*. (d) Estimated by the presented *MPTH*.



Fig. 11. Zoom in of the DROI 2. (a) The noisy image. (b) Estimated by the *STH*. (c) Estimated by the *HTH*. (d) Estimated by the presented *MPTH*.

TABLE V
MSR AND CNR RESULTS OF MRI IMAGES *LIVER* BY THE THREE SCHEMES

Method	DROI 1		DROI 2	
	CNR	MSR	CNR	CNR
Original	2.61	2.61	2.58	2.45
<i>STH</i>	4.08	5.53	4.55	6.12
<i>HTH</i>	4.10	5.58	4.60	6.20
<i>MPTH</i>	4.31	5.98	4.85	6.72

TABLE VI
MSR AND CNR RESULTS OF MRI IMAGES *SPINE* BY THE THREE SCHEMES

Method	DROI 1		DROI 2	
	CNR	MSR	CNR	CNR
Original	2.26	1.79	2.82	2.34
<i>STH</i>	2.96	2.75	3.99	3.86
<i>HTH</i>	2.98	2.78	4.02	3.91
<i>MPTH</i>	3.10	2.91	4.17	4.06

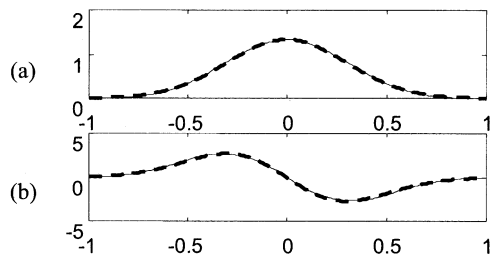


Fig. 12. (a) The smooth function $\theta(x)$ (solid) and a *Gaussian* function (dashed). (b) Wavelet $\psi(x)$ (solid) and the first derivative of the *Gaussian* in (a) (dashed).

structures from noise. The distribution of the products was analyzed and an adaptive threshold was formulated to remove most of the noise. Experiments on the MRI images show that the proposed scheme not only achieves high MSR and CNR measurements but also preserves more edge features.

$$w_j^{x,n} = 1, 2, \dots, 2^{j-1}, w_j^{x,m} = 0, 1, 2, 4, \dots, 2^{j-2}$$

$$w_j^{y,n} = 0, 1, 2, 4, \dots, 2^{j-2}, w_j^{y,m} = 1, 2, \dots, 2^{j-1}$$

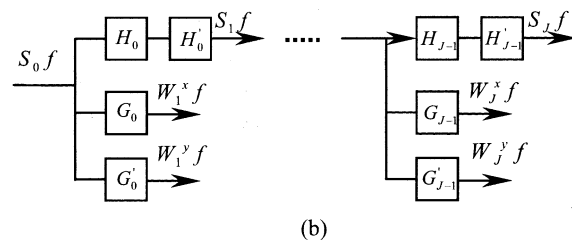
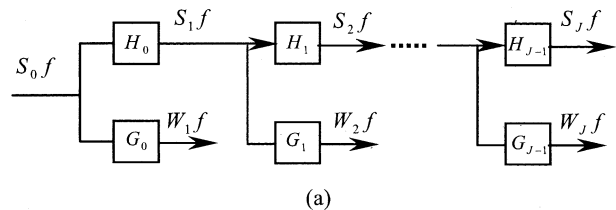


Fig. 13. The discrete decomposition algorithms of (a) 1-D DWT and (b) 2-D DWT. Where filter $H_j(G_j)$ is the 2^j dilation of $H_0(G_0)$ (putting $2^j - 1$ zeros between each of coefficients of $H_0(G_0)$) and $H'_j(G'_j)$ is the transpose of $H_j G_j$.

APPENDIX THE MZ WAVELET AND THE DISCRETE DECOMPOSITION ALGORITHM

Mallat and Zhong [4] defined a class of wavelets that can be used in the implementation of DWT. The FT of the wavelet $\psi(x)$ is

$$\hat{\psi}(\omega) = i\omega \left(\frac{\sin(\frac{\omega}{4})}{\frac{\omega}{4}} \right)^4. \quad (\text{A.1})$$

Therefore, the FT of its associated smooth function $\theta(x)$, the primitive of $\psi(x)$, is

$$\hat{\theta}(\omega) = \left(\frac{\sin(\frac{\omega}{4})}{\frac{\omega}{4}} \right)^4. \quad (\text{A.2})$$

The $\theta(x)$ is a cubic spline whose integral is equal to 1 and $\psi(x)$ is a quadratic spline. In Fig. 12, they are plotted and compared with a Gaussian function and its first derivative. It is noticed that the $\theta(x)$ approximates closely to the *Gaussian* function. The wavelet transform behaves like a Canny edge detector [17].

The discrete decomposition algorithms of 1-D and 2-D DWT are illustrated in Fig. 13. Filter H_j (G_j) is the 2^j scale dilation of H_0 (G_0) (putting $2^j - 1$ zeros between each of the coefficients of H_0 (G_0)). H'_j (G'_j) is the transpose of H_j (G_j). The coefficients of filters H_0 and G_0 are available in [4]. Suppose the input signal $S_0 f$ has N samples, then at each scale 2^j the wavelet coefficients $S_j f$ and $W_j f$ also have N samples. There are at most $\log_2 N$ scales and the complexity of the decomposition algorithm is $O(N \log_2 N)$.

It should be noted that in the discrete implementation, at each scale the wavelet coefficients should be sampled with a constant shift. For a 1-D signal, we denote the discrete sample sequence by

$$dW_j f(n) = W_j f(n + w_j) \quad (\text{A.3})$$

where the shift variable w_j is

$$w_j = 1, 2, \dots, 2^{j-1}. \quad (\text{A.4})$$

For a 2-D image, there are two sample sequences obtained in horizontal and vertical directions

$$\begin{aligned} dW_j^x f(n, m) &= W_j^x f(n + w_j^{x,n}, m + w_j^{x,m}) \\ dW_j^y f(n, m) &= W_j^y f(n + w_j^{y,n}, m + w_j^{y,m}) \end{aligned} \quad (\text{A.5})$$

where the sample shifts are

$$\begin{aligned} w_j^{x,n} &= 1, 2, \dots, 2^{j-1}, w_j^{x,m} = 0, 1, 2, 4, \dots, 2^{j-2} \\ w_j^{y,n} &= 0, 1, 2, 4, \dots, 2^{j-2}, w_j^{y,m} = 1, 2, \dots, 2^{j-1}. \end{aligned} \quad (\text{A.6})$$

REFERENCES

- [1] I. Daubechies, *Ten Lectures on Wavelets*. Philadelphia, PA: SIAM, 1992.
- [2] Y. Meyer, *Wavelets and Operators*. Cambridge, U.K.: Cambridge Univ. Press, 1992.
- [3] S. Mallat and W. L. Hwang, "Singularity detection and processing with wavelets," *IEEE Trans. Inform. Theory*, vol. 32, pp. 617–643, Mar. 1992.

- [4] S. Mallat and S. Zhong, "Characterization of signals from multiscale edges," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 14, pp. 710–732, July 1992.
- [5] M. Vetterli and C. Herley, "Wavelet and filter banks: Theory and design," *IEEE Trans. Signal Processing*, vol. 40, pp. 2207–2232, Sept. 1992.
- [6] J. M. Shapiro, "Embedded image coding using zerostrees of wavelet coefficients," *IEEE Trans. Signal Processing*, vol. 41, pp. 3445–3462, Dec. 1993.
- [7] M. S. Crouse, R. D. Nowak, and R. G. Baraniuk, "Wavelet-based statistical signal processing using hidden Markov models," *IEEE Trans. Signal Processing*, vol. 46, pp. 886–902, Apr. 1998.
- [8] R. D. Nowak, "Wavelet-based Rician noise removal for magnetic resonance," *IEEE Trans. Image Processing*, vol. 8, pp. 1408–1419, Oct. 1999.
- [9] A. Macovski, "Noise in MRI," *Magn. Reson. Med.*, vol. 36, pp. 494–497, 1996.
- [10] H. Gudbjartsson and S. patz, "The Rician distribution of noisy MRI data," *Magn. Reson. Med.*, vol. 34, pp. 910–914, 1995.
- [11] J. C. Wood and K. M. Johnson, "Wavelet packet denoising of magnetic resonance images: Importance of Rician noise at low SNR," *Magn. Reson. Med.*, vol. 41, pp. 631–635, 1999.
- [12] S. Zaroubi and G. Goelman, "Complex denoising of MR data via wavelet analysis: Application for functional MRI," *Magn. Reson. Imag.*, vol. 18, pp. 59–68, 2000.
- [13] M. E. Alexander *et al.*, "A wavelet-based method for improving signal-to-noise ratio and contrast in MR images," *Magn. Reson. Imag.*, vol. 18, pp. 169–180, 2000.
- [14] H. Soltanian-Zadeh, J. P. Windham, and A. E. Yagle, "A multidimensional nonlinear edge-preserving filter for magnetic resonance image restoration," *IEEE Trans. Image Processing*, vol. 4, pp. 147–161, Feb. 1995.
- [15] G. Cincotti, G. Loi, and M. Pappalardo, "Frequency decomposition and compounding of ultrasound medical images with wavelet packets," *IEEE Trans. Med. Imag.*, vol. 20, pp. 764–771, Aug. 2001.
- [16] A. Pizurica, W. Philips, I. Lemahieu, and M. Acheroy, "A versatile wavelet domain noise filtration technique for medical imaging," *IEEE Trans. Med. Imag.*, vol. 22, pp. 323–331, Mar. 2003, to be published.
- [17] J. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-8, pp. 679–697, Nov. 1986.
- [18] Y. Xu, J. B. Weaver, D. M. Healy Jr, and J. Lu, "Wavelet transform domain filters: A spatially selective noise filtration technique," *IEEE Trans. Image Processing*, vol. 3, pp. 747–758, Nov 1994.
- [19] D. L. Donoho and I. M. Johnstone, "Ideal spatial adaptation via wavelet shrinkage," *Biometrika*, vol. 81, pp. 425–455, 1994.
- [20] —, "Adapting to unknown smoothness via wavelet shrinkage," *J. Amer. Statist. Assoc.*, vol. 90, pp. 1200–1224, Dec. 1995.
- [21] R. R. Coifman and D. L. Donoho, "Translation-invariant de-noising," in *Wavelet and Statistics*, A. Antoniadis and G. Oppenheim, Eds. Berlin, Germany: Springer-Verlag, 1995.
- [22] Q. Pan *et al.*, "Two denoising methods by wavelet transform," *IEEE Trans. Signal Processing*, vol. 47, pp. 3401–3406, Dec. 1999.
- [23] S. G. Chang *et al.*, "Adaptive wavelet thresholding for image denoising and compression," *IEEE Trans. Image Processing*, vol. 9, pp. 1532–1546, Sept. 2000.
- [24] B. M. Sadler and A. Swami, "Analysis of multiscale products for step detection and estimation," *IEEE Trans. Inform. Theory*, vol. 45, pp. 1043–1051, Apr. 1999.
- [25] K. S. Miller, *Multidimensional Gaussian Distributions*. New York: Wiley, 1964.
- [26] J. K. Patel and C. B. Read, *Handbook of The Normal Distribution*. New York: Marcel Dekker, 1982.