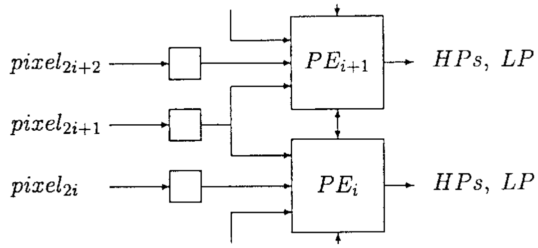

Fig. 3. MPRA implementation.

elements, and the axis *time* is the time dimension. The vectors represent the data dependencies. For example, the computation of level 2 in PE_0 at time four requires the results of the following computations at level 1: PE_1 at time 1, PE_0 at time 2, and PE_{-1} at time 3. Thus, we have three clock cycles to send the result of level 1 from PE_1 to PE_0 and one clock cycle to send the result of level 1 from PE_{-1} to PE_0.

For clarity of the drawings, we include only the dependencies at the first two levels. The dependencies at the following levels have the same slopes but are longer. It should be clear that the same array can compute all the successive levels, as well as the output of the three highpass filters. The nodes with a circle ($\circ$) and a bullet ($\bullet$) are active at level 1, and the nodes with a bullet ($\bullet$) are active at level 2.

The wavelet transform is based on a dilation; thus the present (2-D) example has a decimation along both dimensions. Decimation along the time dimension means that some variables will travel through the array at slower speed. Decimation along the space dimension means that some processing elements will be more active than others. A real implementation may have to redistribute the load amongst neighboring processing elements. Notice that the decimation factor along the time or the space dimension may be different than two.

This implementation requires $W/2$ processing elements (W being the number of lines in the image) and processes an image in $O(W)$ s. The size of the memory required is $O(Wqr)$. Fig. 3 represents the interconnection of two adjacent processing elements. The image is fed from the left, and the wavelet coefficients are produced on the right. Notice that the design requires only a connection between neighboring processing elements.

V. Conclusion

In this paper, we present a generalization of the Mallat pyramid for 2-D wavelets. We introduce a transformation to localize the equations defining the successive levels of the pyramid. We propose a methodology for implementing these wavelet transforms in parallel architectures like systolic arrays. Notice that the methodology is valid whether the wavelet transform is separable or not. We demonstrate the methodology with an example that is a pure systolic architecture based on our previous research in multiphase multirate arrays (MPRA, cf., for example, [6]). This implementation requires $W/2$ processing elements (W being the number of lines in the image) and processes an image in $O(W)$ s. The size of the memory required is $O(Wqr)$.

Acknowledgment

The authors would also like to thank Dr. S. Rajopadhye and the reviewers for helpful comments.

References

[1] M. G. Albanesi and M. Ferretti, “A high speed Haar transform implementation,” *J. Circuits, Syst., Comput.*, vol. 2, no. 3, 1992.
[2] M. Barnsley, *Fractals Everywhere*. New York: Academic, 1988.
[3] K. Gröchenig and W. R. Madych, “Multiresolution analysis, Haar bases, and self-similar tilings of R^n,” *IEEE Trans. Comput.*, Mar. 1992.
[4] J. Kovacevic and M. Vetterli, “Nonseparable multidimensional perfect reconstruction filter banks and wavelet bases for R^n,” *IEEE Trans. Inform. Theory*, vol. 38, pp. 533–555, Mar. 1992.
[5] P. M. Lenders and S. V. Rajopadhye, “Multiphase multirate arrays,” in *Proc. Australas. Comput. Sci. Conf.*, Adelaide, Australia, Jan. 1995.
[6] P. M. Lenders and S. Rajopadhye, “Synthesis of multirate VLSI arrays,” in *Proc. IEEE Int. Conf. Appl.-Specific Array Processors*, Strasbourg, Switzerland, July 1995.
[7] J. Lu, “Parallelizing Mallat algorithm for 2-D wavelet transforms,” *Inform. Process. Lett.*, vol. 45, pp. 255–259, 1993.
[8] S. Mallat, “A theory of multiresolution signal decomposition: The wavelet representation,” *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 11, July 1989.
[9] S. V. Rajopadhye, “Synthesizing systolic arrays with control signals from recurrence equations,” *Distrib. Comput.*, 1989.
[10] M. Vishwanath, “The recursive pyramid algorithm for discrete wavelet transform,” *IEEE Trans. Signal Processing*, vol. 42, pp. 673–676, Mar. 1994.

Denoising by Singularity Detection

Tai-Chiu Hsung, Daniel Pak-Kong Lun, and Wan-Chi Siu

***Abstract*—In this correspondence, a new algorithm for noise reduction using the wavelet transform is proposed. Similar to Mallat’s wavelet transform modulus maxima denoising approach, we estimate the regularity of a signal from the evolution of its wavelet transform coefficients across scales. However, we do not perform maxima detection and processing; therefore, complicated reconstruction is avoided. Instead, the local regularities of a signal are estimated by computing the sum of the modulus of its wavelet coefficients inside the corresponding “cone of influence,” and the coefficients that correspond to the regular part of the signal for reconstruction are selected. The algorithm gives an improved denoising result, as compared with the previous approaches, in terms of mean squared error and visual quality. The new denoising algorithm is also invariant to translation. It does not introduce spurious oscillations and requires very little *a priori* information of the signal or noise. Besides, we extend the method to two dimensions to estimate the regularity of an image by computing the sum of the modulus of its wavelet coefficients inside the so-called “directional cone of influence.” The denoising technique is applied to tomographic image reconstruction, where the improved performance of the new approach can clearly be observed.**

I. Introduction

Denoising by wavelet methods has received much attention recently [1]–[3]. In particular, Mallat *et al.* [1] introduced a denoising technique that makes use of the wavelet transform modulus maxima

Manuscript received June 30, 1997; revised April 18, 1999. This work was supported by the Hong Kong RGC Competitive Bids Research Grant B-Q109. The associate editor coordinating the review of this paper and approving it for publication was Dr. Xiang-Gen Xia.

The authors are with the Centre for Digital Signal Processing for Multimedia Applications, Department of Electronic and Information Engineering, The Hong Kong Polytechnic University, Hong Kong (e-mail: enpklun@polyu.edu.hk).

Publisher Item Identifier S 1053-587X(99)08303-8.

(WTMM) representation. The WTMM representation of a signal records the values and locations of local maxima of its wavelet transform modulus (WTM). They proved that the local Lipschitz exponent of a signal can be estimated by tracing the evolution of its WTMM across scales. From the estimated Lipschitz exponent and with some other *a priori* information of the signal, an effective denoising method can be developed. Although the WTMM-based algorithms give a promising performance in many aspects, the irregular sampling nature of the WTMM complicates the reconstruction process. Furthermore, examples were found [4] to show that the WTMM representation cannot uniquely characterize a signal. It implies that the reconstruction of signal from its WTMM may not be consistently stable. Consequently, many researchers suggested other methods to estimate a signal from its WTMM representation [5], [6]. The discrete time wavelet extrema representation [7] was also proposed as an alternative convex representation for the WTMM. However, the reconstruction method is still iterative, which leads to the high computational complexity of these kinds of approaches.

On the other hand, Donoho [2], [3] suggested another stream of denoising technique by performing hard thresholding or shrinkage on the orthogonal wavelet transform coefficients of a signal. Due to the vanishing moment property and the compact support of wavelets, most signal energy, after wavelet transform, is supposed to be clustered in a few wavelet coefficients, whereas noises do not. The thresholding or shrinkage on the wavelet coefficients with a proper threshold [8] can then significantly reduce noise. This operation guarantees with high probability that the denoised signal is at least as smooth as the input noisy signal. However, the denoised signal may contain spurious oscillations due to the translation-variant property of the decimating wavelet transform. Therefore, many variants of the wavelet shrinkage techniques were developed. They include the "cycle spinning" approach [9] and the approaches using undecimated wavelet basis [10] and near shift-invariant wavelet bases [11]. Nevertheless, not all of these approaches guarantee that edges can be preserved in denoising due to the neglect of interscale information. This artifact becomes important when the denoising technique is applied to 2-D images since human perception is sensitive to image edges.

In this correspondence, a new algorithm for noise reduction using the wavelet transform is proposed. The new algorithm can be viewed as a combination of Mallat and Donoho's denoising approaches. As in the wavelet thresholding approaches, the new algorithm selects the desired wavelet transform coefficients for reconstruction based on a threshold. However, the threshold is determined based on the estimated Lipschitz exponents of the signal, as in the WTMM approach. Although Mallat's algorithm has the drawback of high computational complexity in reconstruction, the new approach has the advantage in that it avoids the complicated reconstruction process. The new algorithm performs better than the original wavelet shrinkage approach [2] in that it preserves the edges of a noisy signal. In the next section, we state and describe some important theories for wavelet denoising by singularity detection. Then, we develop a new denoising algorithm using the sum of the WTM inside the "cone of influence" to estimate the regularity of a signal. We further develop a new two-dimensional (2-D) denoising algorithm using the sum of the WTM inside the so-called "directional cone of influence." Finally, we give the performance of the new denoising algorithms.

II. Denoising Algorithm Using the WTMM

The discrete dyadic wavelet transform [1], [12], [13] of a one-dimensional (1-D) discrete signal $f \in L^2(R)$ is defined as $\{S_{2^J} f, (W_{2^j} f)_{1 \le j \le J}\}$. The components are obtained by the convolutions of $f(x)$ with the scaling function and the dilated wavelets $S_{2^J} f = f * \phi_{2^J}(x)$, $W_{2^j} f(x) = f * \psi_{2^j}(x)$. As indicated in [1], we can estimate the local Lipschitz exponent for a particular point of a signal by the following theorem.

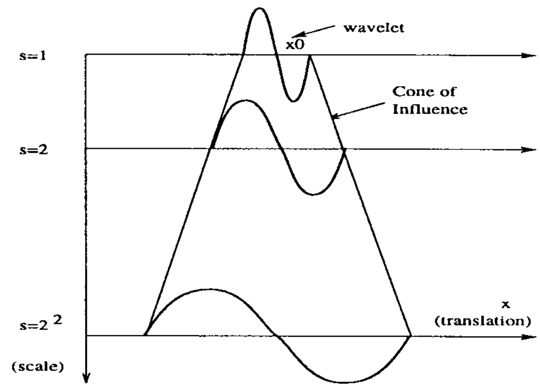

Fig. 1. "Cone of influence," which is the support of the wavelet function in different scales.

Theorem 1: If $f(x)$ is Lipschitz α at x_0, then there exists a constant A such that for all points x in the neighborhood of x_0 and any scale s

$$|W_s f(x)| \le A(s^\alpha + |x - x_0|^\alpha). \tag{1}$$

Conversely, $f(x)$ is Lipschitz α at x_0 if two conditions hold:

- There exist some $\epsilon > 0$ and a constant A such that for all x in the neighborhood of x_0 and any scale s
$$|W_s f(x)| \le A s^\epsilon. \tag{2}$$
- There exists a constant B such that for all x in the neighborhood of x_0 and any scale s
$$|W_s f(x)| \le B\left(s^\alpha + \frac{|x - x_0|^\alpha}{|\log |x - x_o||}\right). \tag{3}$$

Equations (1) and (3) imply that $|W_s f(x)| \le O(s^\alpha)$ inside a cone $|x - x_0| \le Ks$ [1], where K is the support of the mother wavelet. This cone is the so-called "cone of influence" (COI), as shown in Fig. 1. Mallat and Hwang furthered [1, Th. 1] and proposed to estimate the Lipschitz exponent of a singularity by tracing its WTMM curves across scales inside the COI. They showed that the local regularity of certain types of nonisolated singularities in the signals can be characterized by using the WTMM. They also showed that the decay of the expected WTMM value of a wide sense stationary white noise across scales is proportional to $1/2^j$, where $s = 2^j$. This means that the WTMM curves of noises are expected to decay across scales at least at a rate of $1/2^j$ or even not propagate to coarser scales. This is not the case for regular signals and edges. Since signal edges possess zero Lipschitz exponents and regular signals possess positive Lipschitz exponents, the corresponding WTMM will be the same, if it does not increase, when scale increases. Consequently, in the application of denoising, we can remove noises of a signal by removing all the WTMM of which the amplitude decreases when scale increases. $|W_{2^j} f(x_{i-1})| \le |W_{2^j} f(x_i)|$, and $|W_{2^j} f(x_i)| \ge |W_{2^j} f(x_{i+1})|$. We trace the maxima curves inside the COI in scale-space to estimate the Lipschitz exponents of the signal singularities. From the estimated Lipschitz exponents, we can process the maxima curves in scale space to obtain the denoised WTMM. Note that there may be many errors and ambiguities in tracing the maxima curves in scale space. They may affect the accuracies of the estimated Lipschitz exponents. This situation appears more significant if the singularities are not isolated. That is, the COI's of these singularities have common support. In this case, we may falsely remove maxima that correspond to the desired singularities or retain maxima that correspond to the undesired singularities.

Besides the processing of maxima curves, another burden in WTMM processing is the irregularly located maxima that complicate the reconstruction process [1], [5]–[7]. Traditionally, the projections onto convex sets (POCS) method [14] was adopted to reconstruct the processed WTMM. The iterative nature of this method introduces much difficulty for the denoising algorithm to be applied to real-time applications.

A. Wavelet Transform Modulus Sum under the Cone of Influence

In order to look for a simpler approach, we move a step back from the maxima processing. For the wavelet coefficients of a signal in each scale, we compute the integral of the modulus of these wavelet coefficients inside the COI. Let us define an operator N, dubbed the wavelet transform modulus sum (WTMS), such that for all point x_0 of a function f

$$N_s f(x_0) = \int_{|x-x_0|\le Ks} |W_s f(x)|\, dx \tag{4}$$

where s is the scale, and K is the support of the mother wavelet function, which is a constant. From Theorem 1

$$\begin{aligned} N_s f(x_0) &\le \int_{|x-x_0|\le Ks} A(s^\alpha + |x-x_0|^\alpha)\, dx \\ &\le 2A\left(K + \frac{K^{\alpha+1}}{\alpha+1}\right)s^{\alpha+1} \le A' s^{\alpha+1} \end{aligned} \tag{5}$$

where A' is a constant. As shown in (5), we can estimate the Lipschitz exponent α from the upper bound of the slope of $\log(N_s f)$ instead of that in the previous approach, where we need to fit α from a set of neighboring coefficients in scale-space or trace the wavelet maxima across scales. It avoids the ambiguous operations, such as tracing of maxima curves in scale space, as in the previous approaches. Indeed, the main difference of using $N_s f(x)$ instead of the maxima of $W_s f(x)$, as in [1], is that the processing of $N_s f(x)$ is over the regularly located wavelet coefficients, whereas the processing of wavelet maxima is over irregularly located maximum points of the wavelet coefficients. The current approach reduces the complexity in realization.

B. Interscale Ratio of WTMS

For a particular point x_o of a function f, we obtain, by using (4), the function $N_s f(x_o)$, which is the sum of the modulus of wavelet coefficients of this point in scale s inside the COI. It is mentioned above that the Lipschitz exponent α for a particular point x_o of the function f can be estimated based on the function $N_s f(x_o)$. However, in the application of denoising, we need not directly estimate the Lipschitz exponent. It is known that [1] the Lipschitz exponent α of Gaussian noise usually possesses negative value when measuring by using the wavelet transform approach with the wavelet of one vanishing moment. In the case that $\alpha \le -1$

$$N_{2^{j+1}} f(x_o) \le N_{2^j} f(x_o) \quad \text{for } 1 \le j < J$$

where J is the total number of scales. It implies that for a strong irregular point that has $\alpha \le -1$, we can easily detect it by measuring the interscale ratio of $N_s f(x_o)$ such that

$$\frac{N_{2^{j+1}} f(x_o)}{N_{2^j} f(x_o)} = 2^{\alpha+1} \le 1 \quad \text{for } 1 \le j < J \quad \text{and} \quad \alpha \le -1.$$

That is, the function $N_s f(x_o)$ will decrease or remain the same as the scale increases. If the point x_o corresponds to the edge or the regular part of the function f, it is known that [1] by using the same wavelet as before, the Lipschitz exponent α at this point is greater than or equal to 0. This point can again be easily detected by measuring the interscale ratio of $N_s f(x_o)$ such that

$$\frac{N_{2^{j+1}} f(x_0)}{N_{2^j} f(x_0)} = 2^{\alpha+1} \ge 2 \quad \text{for } 1 \le j < J \quad \text{and} \quad \alpha \ge 0. \tag{6}$$

That is, the function $N_s f(x_o)$ will increase at least doubly as scale increases. By selecting the wavelet coefficients that fulfill the "interscale ratio" condition, as stated in (6), we can effectively remove noise while the edges and the regular part of the signal can be preserved.

C. Interscale Difference of WTMS

The use of the interscale ratio method provides a simple means to select the wavelet coefficients that correspond to the regular parts of the signal. Furthermore, it has the advantage that it requires no *a priori* information about the noise or signal. In other words, its performance will not be affected due to the variation of noise or signal levels. Nevertheless, due to the error generated during the estimation of $N_s f(x)$ as well as the alias that may introduce from the COI of a nearby singularity, it is observed that some small irregular signal will have its wavelet coefficients fulfill the criterion defined in (6). This effect is more obvious for those irregular signals that have $-1 < \alpha < 0$. Since, for these signals, the sum of the modulus of their wavelet coefficients will also increase as the scale increases. The only difference of them from a regular signal is that they have slower increasing rates. However, the introduction of the errors, as mentioned above, can increase their rates of increment across scales to enable them to falsely fulfill the criterion as stated in (6). To solve this problem, we consider the interscale difference condition, as shown in (7)

$$N_{2^{j+1}} f(x_0) - N_{2^j} f(x_0) > \gamma \tag{7}$$

where γ is a threshold. The edge and regular part of the signal can be extracted out by using (7) with appropriately selected threshold $\gamma \ge A' 2^j$ since

$$N_{2^{j+1}} f(x_0) - N_{2^j} f(x_0) = (2^{\alpha+1} - 1)A'(2^j)^{\alpha+1} \ge A' 2^j \tag{8}$$

for $\alpha \ge 0$. Equation (8) shows that even if the wavelet coefficients of a signal can fulfill the criterion in (6), their magnitude must also be great enough to fulfill the criterion in (8). In other words, small irregular signals will not fulfill the criterion in (8), and their corresponding wavelet coefficients will be rejected. However, the constant A' depends on the amplitude of the signal itself and varies among signals. If the value of threshold γ is set too high, the wavelet coefficients at x_0 would be rejected even if it is regular. Fortunately, since the introduction of the interscale difference criterion is to remove the small irregular signal, the threshold γ should always be small; hence, the probability of rejecting the regular part of the signal is also small. For the case that the value of γ is not comparable to the amplitude of the signal at any location, i.e., $\gamma \ll A' 2^j$ or $\gamma = 0$, the condition as stated in (7) becomes the original interscale ratio condition, as stated in (6), since all data that satisfy (6) will satisfy (7). Therefore, the threshold γ can be considered to be a tuning parameter to balance the denoising of the irregular signal and the preservation of the regular signal.

In practice, certain tolerance is allowed in using (6) and (7) to select the required wavelet coefficients. It is particularly important

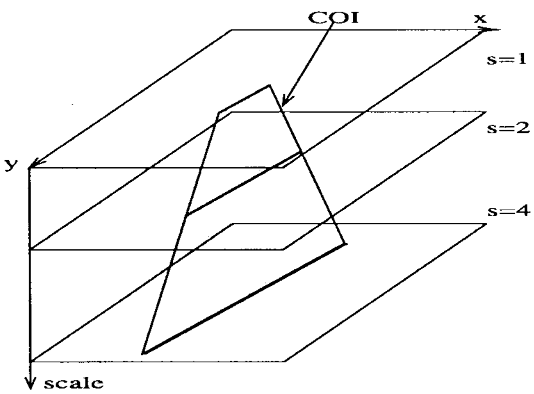

Fig. 2. "Directional cone of influence," which is support of wavelets in different scales with direction indicated by $A_s f(x_o, y_0)$.

for some classes of signals with known regularity range. The resulted wavelet coefficients using the above criteria can be considered to be the ones that correspond to essentially the regular parts of a signal. The denoised signal is then reconstructed from the selected wavelet coefficients using the conventional inverse wavelet transform. It should be noted that the selected wavelet coefficients using the interscale ratio and the interscale difference conditions, essentially from a complete set of wavelet coefficients, correspond to the regular parts of the signal. Hence, we can use a simple inverse wavelet transform for the reconstruction. However, for the original approach in [1], only the WTMM, which belongs to an incomplete set of the required wavelet coefficients, result from the algorithm. Hence, further computation is required to estimate the missing wavelet coefficients from that incomplete set, thus requiring a long computation time.

To summarize, we consider the interscale information of a signal to select the wavelet coefficients that most likely correspond to the regular parts of the signal. The process of singularity detection is local and requires very little *a priori* information of the signal to determine the threshold in using the interscale difference condition. The threshold γ is supposed to be kept small, and in the case that $\gamma = 0$, it degenerates to the case of using an interscale ratio condition. Other merits of the proposed denoising algorithm include the preservation of signal edges and the translation-invariant property. Indeed, it is not easy for other wavelet denoising approaches, such as the wavelet thresholding or shrinkage approach using orthogonal wavelet bases, to preserve the signal edges when denoising since, in those cases, it is difficult to explicitly reveal the edges of a signal from the corresponding wavelet coefficients, especially using translation-variant wavelet bases. From the above analysis, we propose a new denoising algorithm, as follows.

Denoising Algorithm:

1) $\forall j$, where $1 \leq j < J$, compute $W_{2^j} f$ and $N_{2^j} f$;
2) Select $W_{2^j} f(x_0)$ if $N_{2^{j+1}} f(x_0)/N_{2^j} f(x_0) > 2$ and $N_{2^{j+1}} f(x_0) - N_{2^j} f(x_0) \geq \gamma$, where γ is a threshold value;
3) Repeat step 2 until all the coefficients in the selected levels have been processed.
4) Reconstruct signal from the selected wavelet coefficients using the inverse wavelet transform.

We show the experimental results in Fig. 3, where the test signal is corrupted by both the shot noise and the Gaussian noise of different noise levels introduced to different parts of the signal. The test signal is divided into ten regions and contaminated by Gaussian noise with ten different noise variances ranging from 0.1 to 0.2. Besides, the signal is also contaminated by the shot noise. These results demonstrate the robustness among different denoising algorithms for nonstationary noise without oracle. For comparison, we implemented the wavelet shrinkage [Fig. 3(b)] using Daubechies' wavelet with four vanishing moments. The threshold value $\lambda\epsilon$ is selected to be 1.860ϵ, as suggested in [2], where ϵ is the noise variance that is supposed to be known by an oracle. The error measure we used is $\text{MSE} = \sum_x (\hat{f}(x) - f_o(x))^2/N$, where $\hat{f}$ is the estimated signal, and f_o is the original signal ($\sum_x f_o(x)^2/N = 1.200\,389$). For the wavelet shrinkage approaches, all wavelet coefficients are shrunk using the same threshold, and therefore, some regular parts of the signal are falsely removed by the algorithm. Furthermore, shot noise with large magnitude cannot be removed (for example, the shot noise near abscissa 50). However, there is no such problem for the proposed singularity detection method, as shown in Fig. 3(c). This is because the proposed method considers the local regularity of the signal; hence, only the local information is used. Furthermore, the algorithm requires no *a priori* information of the signal or noise. Its performance will not be affected by the variation of noise variance within the signal.

III. Two-Dimensional Denoising Algorithm

The denoising algorithm is extended for two-dimensional (2-D) signals in this section. The approach is basically similar to the 1-D case. However, we extend the concept of COI to a 2-D form to be the so-called "directional cone of influence" (DCOI). First, let us define the discrete dyadic wavelet transform [1], [12] of a discrete image f to be $\{S_{2^J} f, (W^1_{2^j} f)_{1 \leq j \leq J}, (W^2_{2^j} f)_{1 \leq j \leq J}\}$, where $f \in L^2(R^2)$. The components are obtained by the convolutions of $f(x,y)$ with the scaling function and the dilated wavelets $S_{2^J} f = f * \phi_{2^J}(x,y)$, $W^1_{2^j} f(x,y) = f * \psi^1_{2^j}(x,y)$, $W^2_{2^j} f(x,y) = f * \psi^2_{2^j}(x,y)$. The wavelets are designed to be the partial derivatives of a smooth function along the x and y directions, respectively. That is, $\psi^1(x,y) = \partial\Phi(x,y)/\partial x$ and $\psi^2(x,y) = \partial\Phi(x,y)/\partial y$. In addition, denote the modulus of the wavelet transform $M_{2^j} f(x,y) = \sqrt{|W^1_{2^j} f(x,y)|^2 + |W^2_{2^j} f(x,y)|^2}$ and the phase $A_{2^j} f(x,y) = \arctan(W^2_{2^j} f(x,y)/W^1_{2^j} f(x,y))$. They indicate the magnitude and orientation of the gradient vector of the wavelet coefficient at a particular point (x,y). Since the orientation of the gradient vector of the wavelet coefficients indicates the direction where the maximum local variation of a signal is found [1], we only need to measure the Lipschitz exponent in that direction in order to identify the singularity of $f(x,y)$. The COI derived from that direction is the so-called DCOI, as illustrated in Fig. 2. In that direction, the characterization of the Lipschitz exponent in Theorem 1 concerning the asymptotic decay of wavelet coefficients across scales is also valid. More precisely, the Lipschitz exponent α at a point (x_0, y_0) of a 2-D function f in a particular direction is related to the modulus of the wavelet coefficients of this function in scales s by

$$|M_s f(x,y)| \leq B s^\alpha$$

where B is a constant. Instead of directly fitting the Lipschitz exponent from a set of neighboring coefficients in scale-space or tracing the maxima curves across scales as indicated in [1], we compute the integral of the modulus of the wavelet coefficients inside the DCOI. We define an operator N, which is dubbed as the wavelet transform modulus sum (WTMS) inside the directional COI of a function such that

$$N_s f(x_0, y_0) = \int_{(x,y) \in D_s} M_s f(x,y)\, dx\, dy \tag{9}$$

where $D_s = \{(x,y) : (x-x_0)^2 + (y-y_0)^2 \leq K s^2,\ (y-y_0)/(x-x_0) = \tan(A_{2^j} f(x_0,y_0))\}$ is the DCOI. We implement the line integral of (9) by linear interpolation since not all wavelet coefficients

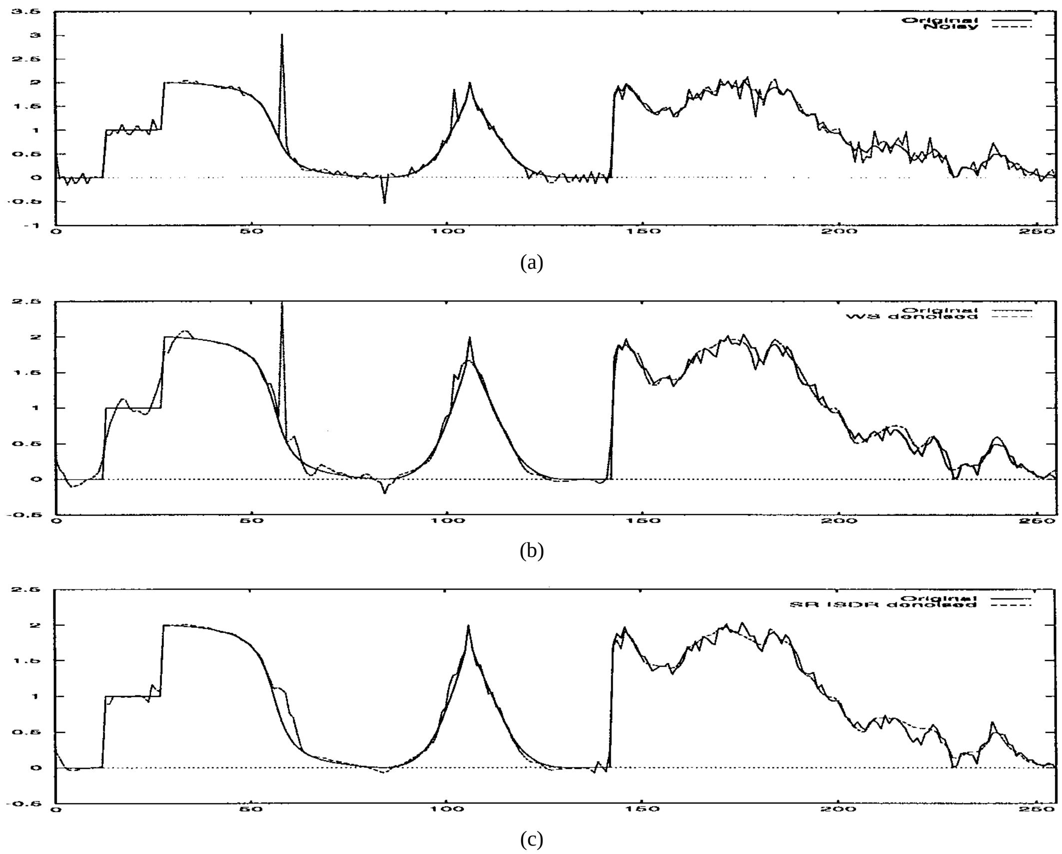

Fig. 3. Comparison of the traditional wavelet denoising approaches and the new denoising algorithm. (a) Noisy and original signal. MSE $= 0.036\,909$. (b) Wavelet shrinkage result and original signal. MSE $= 0.023\,712$. (c) Result of singularity detection by using the interscale ratio and the interscale differences as compared with the original signal. MSE $= 0.008\,133$.

lie on the direction indicated by $A_s f$. As the DCOI is just the COI in a particular direction on a 2-D plane, we can again make use of Theorem 1 to rewrite (9) as

$$N_s f(x_0, y_0) \leq B' s^{\alpha+1}. \tag{10}$$

The proof of (10) is similar to that for (5). Therefore, the Lipschitz exponent α can be estimated from the upper bound of the slope of $\log(N_s f)$, as in the 1-D case. In the application of denoising, we do not need to directly estimate the Lipschitz exponent α. As in the 1-D case, we perform the wavelet transform on the noisy signal and select the wavelet coefficients that give positive Lipschitz exponents. The wavelet that we used is the first derivative of a smoothing function that has one vanishing moment. This gives rise to the interscale ratio condition, as in similar in the 1-D case

$$N_{2^{j+1}} f(x_0, y_0)/N_{2^j} f(x_0, y_0) > 2. \tag{11}$$

As in the 1-D case, small irregular signals may falsely fulfill the condition, as stated in (11), due to the error generated in measuring the directional sum. Therefore, we also introduce the interscale difference condition to reject the small irregular signal

$$N_{2^{j+1}} f(x_0, y_0) - N_{2^j} f(x_0, y_0) > \gamma \tag{12}$$

where γ is a threshold. Basically, the applications of the 2-D interscale ratio condition and the interscale difference condition are the same as in the 1-D case, but the WTMS in each scale is obtained only in the direction indicated by $A_s f(x, y)$ instead of the whole three-dimensional (3-D) COI. This significantly reduces the complexity in computing the WTMS. As mentioned in the 1-D case, one of the major advantages of the proposed singularity detection denoising algorithm is that it does not require a complicated reconstruction process for the selected wavelet coefficients. It is also true for the 2-D case. The proposed approach requires only a simple inverse wavelet transform to reconstruct the selected wavelet coefficients. The saving of computation time, as compared with that in [1], can be up to a few orders of magnitude. To conclude, the proposed 2-D denoising algorithm can be summarized as follows.

2-D Denoising Algorithm:

1) $\forall j$ where $1 \leq j < J$, compute $W_{2^j} f$ and $N_{2^j} f$ as described in (9).
2) Select $W^1_{2^j} f(x_0, y_0), W^2_{2^j} f(x_0, y_0)$ if $N_{2^{j+1}} f(x_0, y_0) / N_{2^j} f(x_0, y_0) \geq 2$ and $N_{2^{j+1}} f(x_0, y_0) - N_{2^j} f(x_0, y_0) \geq \gamma$, where γ is a threshold value;
3) Repeat step 2 until all the coefficients in the selected levels have been processed.
4) Reconstruct the image from the selected wavelet coefficients using the inverse wavelet transform.

To demonstrate the results of denoising for 2-D images, we apply the 2-D denoising algorithm to the application of tomographic image reconstruction. The algorithm is applied to the sinogram [15] (the projections of an image). For comparison, different denoising techniques are applied to the noisy sinogram with a 128 $\times$ 128 projection set. Fig. 4(b) shows the original sinogram and the image reconstructed from the projections using the filtered back-projection (FBP) algorithm [15]. The sinogram is then contaminated with zero mean Gaussian noise of variance 0.5. For the case where the noise variance is 0.5, Fig. 4(a) shows the noisy sinogram with MSE equal to 0.25 and the reconstructed image with MSE equal to 1.59. The signal power for the original sinogram and its reconstruction

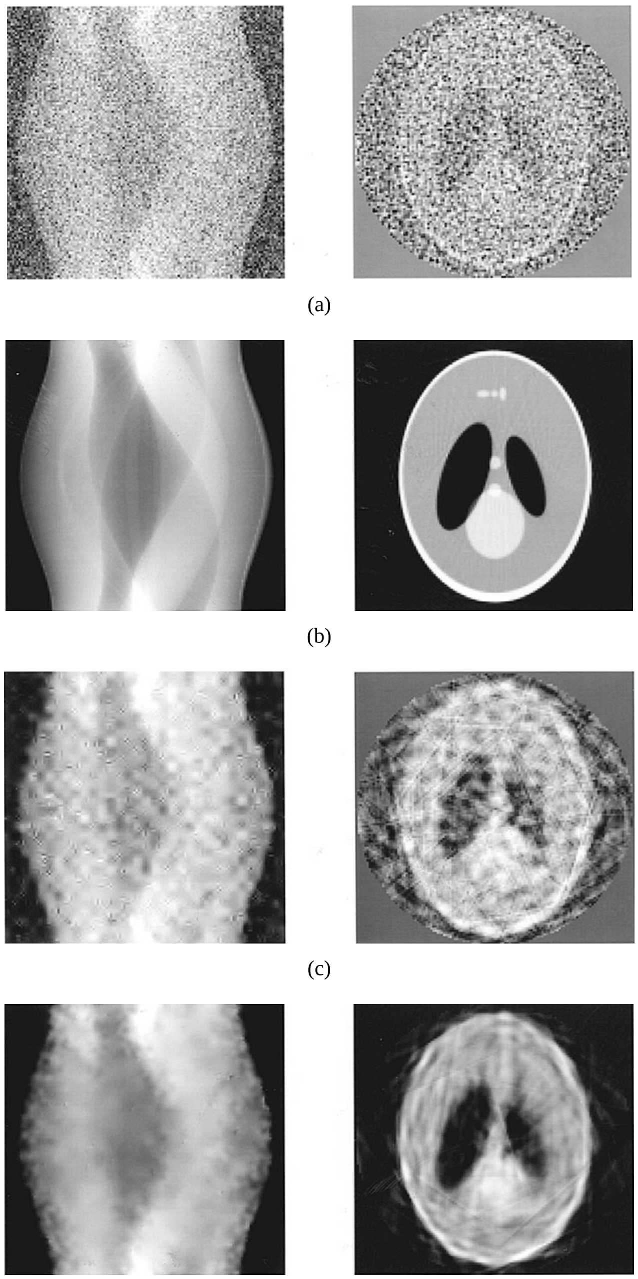

Fig. 4. Comparision of orthogonal wavelet thresholding and the proposed denoising algorithm in the reconstruction of image from its sinogram with Gaussian noise of zero mean and 0.5 variance. (a) MSE $= 0.248\,665$ MSE $= 1.590\,063$ Noisy sinogram and FBP image. (b) Original sinogram and FBP image. (c) MSE $= 0.021\,426$ MSE $= 0.121\,357$ Denoised sinogram and FBP image using the wavelet shrinkage. (d) MSE $= 0.007\,339$ MSE $= 0.041\,172$ Denoised sinogram and FBP image using the singularity detection with interscale difference.

are $\sum_{x,y=0}^{N-1} \check{f}(x,y)^2/N^2 = 1.2902$ and $\sum_{x,y=0}^{N-1} f(x,y)^2/N^2 = 0.5950$, respectively, for $N = 128$. Fig. 4(c) shows the denoised result using the wavelet shrinkage approach, with universal threshold $\lambda\epsilon$, where $\lambda = 1.860$, as suggested in [2]. Fig. 4(d) shows the denoised result using the proposed approach. The reconstructed image shows that the result using the new algorithm is better than that of the wavelet shrinkage approach in terms of MSE and visual quality. The improvement is due to the fact that the proposed approach considers the information of the wavelet coefficients evolved across scales.

IV. Conclusion

In this correspondence, we proposed a new algorithm for noise reduction using the wavelet transform modulus sum (WTMS). We suggest the use of both the interscale ratio and the interscale difference conditions of the WTMS to select the required wavelet coefficients. Merits of the singularity detection denoising algorithm include edges preservation and translation invariance. It does not introduce spurious oscillations and only requires very little *a priori* information of the signal. In addition, we extend the method to two-dimensions for image denoising. By computing the WTMS of the noisy image inside the corresponding “directional cone of influence” and applying the interscale ratio and the interscale difference conditions, the irregular parts of the image are readily identified. The denoising technique is applied to tomographic image reconstruction, where the improved performance of the new approach is clearly shown.

References

[1] S. Mallat and W. L. Hwang, “Singularity detection and processing with wavelets,” *IEEE Trans. Inform. Theory*, vol. 38, pp. 617–643, Mar. 1992.
[2] D. L. Donoho and I. M. Johnstone, “Ideal spatial adaptation by wavelet shrinkage,” *Biometrika*, vol. 81, no. 3, pp. 425–455, 1994.
[3] D. L. Donoho, “De-noising by soft-thresholding,” *IEEE Trans. Inform. Theory*, vol. 41, pp. 613–627, May 1995.
[4] Y. Meyer, “Un contre-example á la conjecture de marr et á celle de S. Mallat,” preprint, 1991.
[5] A. Liew and D. T. Nguyen, “Uniqueness issue of wavelet transform modulus maxima representation and a least squares reconstruction algorithm,” *Electron. Lett.*, vol. 31, no. 20, pp. 1735–1736, Sept. 28, 1995.
[6] ——, “Modulus maxima using nonexpansive projections,” *Electron. Lett.*, vol. 31, no. 13, pp. 1038–1039, June 22, 1995.
[7] Z. Cvetković and M. Vetterli, “Discrete-time wavelet extrema representation: Design and consistent reconstruction,” *IEEE Trans. Signal Processing*, vol. 43, pp. 681–693, Mar. 1995.
[8] D. L. Donoho and I. M. Johnstone, “Threshold selection for wavelet shrinkage of noisy data,” in *Proc. 16th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.—Eng. Adv.: New Opportunities Biomed. Eng.*, 1994, vol. 1, pp. A24–A25.
[9] R. R. Coifman and D. L. Donoho, “Translation-invariant denoising,” in *Wavelets and Statistics*. New York: Springer-Verlag, 1995, pp. 125–150.
[10] M. Lang, H. Guo, J. E. Odegard, C. S. Burrus, and R. O. Wells, “Noise reduction using undecimated discrete wavelet transform,” *IEEE Signal Processing Lett.*, vol. 3, pp. 10–12, Jan. 1996.
[11] Y. Hui, C. W. Kok, and T. Q. Nguyen, “Wavelet shrinkage denoising using paraunitary shift-invariant filter banks,” in *Proc. 1997 IEEE Int. Symp. Circuits Syst.*, 1997, vol. 1, pp. 185–188.
[12] S. 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. Mallat, “Zero-crossings of a wavelet transform,” *IEEE Trans. Inform. Theory*, vol. 37, pp. 1019–1033, July 1991.
[14] D. C. Youla and H. Webb, “Image restoration by the method of convex projections,” *IEEE Trans. Med. Imag.*, vol. MI-1, pp. 81–101, Oct. 1982.
[15] S. R. Deans, *The Radon Transform and Some of Its Applications*. New York: Wiley, 1983.