

Wavelet-Based Multiscale Anisotropic Diffusion with Adaptive Statistical Analysis for Image Restoration

Junmei Zhong, Huifang Sun

TR2008-067 November 2008

Abstract

The anisotropic diffusion techniques are in general efficient to preserve image edges when they are used to reduce noise. However, they are not very effective to denoise those images that are corrupted by a high level of noise mainly for the lack of a reliable edge-stopping criterion in the partial differential equation (PDE). In this paper, a new algorithm is developed to tackle this problem. The main contribution of this paper is in the construction of a new regularization method for the PDE by using the overcompleted dyadic wavelet transform (DWT). It proposes to perform anisotropic diffusion in the more stationary DWT domain rather than directly in the raw noisy image domain. In the DWT domain, since noise tends to decrease as the scale increases, at each scale, noise has less influence on the PDE than that in the raw noisy image domain. As a result, the edge-stopping criterion and other partial derivative measurements in the PDE become more reliable. Furthermore, there is no need to do Gaussian smoothing or any other smoothing operations. Experiment results show that the proposed algorithm can significantly reduce noise while preserving image edges.

IEEE

This work may not be copied or reproduced in whole or in part for any commercial purpose. Permission to copy in whole or in part without payment of fee is granted for nonprofit educational and research purposes provided that all such whole or partial copies include the following: a notice that such copying is by permission of Mitsubishi Electric Research Laboratories, Inc.; an acknowledgment of the authors and individual contributions to the work; and all applicable portions of the copyright notice. Copying, reproduction, or republishing for any other purpose shall require a license with payment of fee to Mitsubishi Electric Research Laboratories, Inc. All rights reserved.

Wavelet-Based Multiscale Anisotropic Diffusion with Adaptive Statistical Analysis for Image Restoration

Junmei Zhong^{*}, IEEE Member, Huifang Sun⁺, IEEE Fellow
^{*}: Dept. of Radiology, University of Cincinnati, Cincinnati, Ohio 45229
⁺: Technology Laboratory, Mitsubishi Electric Research Laboratories
 201 Broadway, Cambridge, MA 02139

Abstract

In general, the anisotropic diffusion techniques are efficient to preserve image edges when they are used for reducing noise. However, they are not very efficient to reduce noise for those images that are corrupted by a high level of noise mainly for the lack of a reliable edge-stopping criterion in the partial differential equation (PDE). In this paper, a new algorithm is developed to address this problem. The contribution of this paper is on the construction of a new regularization method for the PDE using wavelet transform without doing Gaussian smoothing or any other smoothing. As a result, the edge-stopping criterion of the PDE and other gradient measurements are more reliable, making the anisotropic diffusion more efficient for noise reduction. Experimental results have shown that the proposed algorithm can significantly reduce noise while preserving image edges.

Key words: Wavelet transform, scale-space, multiscale anisotropic diffusion, adaptive statistical analysis, image restoration.

I. Introduction

For image denoising, the main challenge is how to preserve the information-bearing structures such as edges and object boundaries to get satisfactory visual quality when improving the signal-to-noise-ratio (SNR). Edge-preserving image denoising has become a very intensive research topic. Traditional Gaussian smoothing is not efficient for preserving edges since the Gaussian kernel is symmetric and orientation-insensitive, resulting in blurring artifact for edges. In the past two decades, the nonlinear anisotropic diffusion model [1] and its variants have been widely used for general image denoising [2-7] and medical imaging [8-11] since it can adaptively encourage the intra-region smoothing while inhibiting the inter-region diffusion for achieving simultaneous noise reduction and edge preservation. However, we find out that the anisotropic diffusion model is very sensitive to noise and is inefficient to denoise those images with low SNR. This is mainly due to the fact that in the partial differential equation (PDE), the anisotropic diffusion coefficient is a non-negative monotonically decreasing function of the image gradient magnitude, but the gradient is very sensitive to noise, making the anisotropic diffusion coefficient non-reliable. In addition, the PDE contains some other partial derivatives and their numerical calculations are also very sensitive to noise.

In this paper, we develop a new algorithm of wavelet based multiscale anisotropic diffusion with adaptive statistical analysis, which can take advantage of edge-preserving property of the anisotropic diffusion model while circumventing its noise sensitivity problem. The main idea of this algorithm is to reduce the influence of noise on the PDE model. For achieving this goal, we use the dyadic wavelet transform (DWT) [12-13] to construct a linear scale-space for the noisy image. Due to the smoothing functionality of the scaling function, the wavelet-based multiscale representation of the noisy image is much more stationary than the raw noisy image. Noise is mostly located in the finest scale and tends to decrease as the scale increases. Afterwards, we perform the minimum mean squared error (MMSE)-based filtering on the finest scale, making the wavelet-based scale-space even more stationary. Finally, we perform the anisotropic diffusion on the stationary scale-space rather than on the rough noisy image domain. Furthermore, the stationary scale-space

makes it possible to optimize the anisotropic diffusion model by removing the regularization component such as Gaussian smoothing or other regularization methods as done in the previous work [2-3]. Comparative studies demonstrate that the proposed algorithm can significantly improve SNR while preserving edges.

The rest of this paper is organized as follows: Section II reviews the related work about anisotropic diffusion. Section III presents the algorithm of wavelet-based multiscale anisotropic diffusion and adaptive statistical analysis. The experimental results are demonstrated in Section IV and conclusion is made in Section V.

II. Related Work about Anisotropic Diffusion

The nonlinear PDE based anisotropic diffusion model proposed by Perona and Malik is to improve the isotropic diffusion [1]. In the continuous domain, consider the isotropic diffusion equation (the heat equation),

$$\frac{\partial I(x, y, t)}{\partial t} = \text{div}(\nabla I) \quad (1)$$

where div denotes the divergence operator. Smoothing an image according to the isotropic diffusion equation in the following way is equivalent to filtering the image with a Gaussian filter

$$I(x, y, t + 1) = I(x, y, t) + \text{div}(\nabla I) \quad (2)$$

The disadvantage of the Gaussian smoothing is that it is symmetric and orientation-insensitive, which would result in blurred edges and reduced resolution. In the anisotropic diffusion, the isotropic diffusion equation is modified into

$$\frac{\partial I(x, y, t)}{\partial t} = \text{div}(g(|\nabla I|)\nabla I) \quad (3)$$

where $|\nabla I|$ is the image gradient magnitude for edge detection, and $g(|\nabla I|)$ is the anisotropic diffusion coefficient, which is a non-negative monotonically decreasing function of the image gradient magnitude. The function of anisotropic diffusion coefficient is defined in such a way that when $|\nabla I| \rightarrow \infty$, $g(|\nabla I|) \rightarrow 0$; and when $|\nabla I| = 0$, $g(|\nabla I|) \rightarrow 1$. Thus the smoothing is only encouraged within homogeneous regions and is prohibited across object boundaries and edges. In terms of the anisotropic diffusion, the image smoothing is defined as following [1]:

$$I(x, y, t + 1) = I(x, y, t) + \text{div}(g(|\nabla I|)\nabla I). \quad (4)$$

The iterative anisotropic diffusion process can be discretized as:

$$I(x, y, t + 1) = I(x, y, t) + \frac{\lambda}{|\eta_{(x,y)}|} \sum_{(p,q) \in \eta_{(x,y)}} g(|\nabla I_{(p,q)}^{(x,y)}|) \nabla I_{(p,q)}^{(x,y)} \quad (5)$$

where the original image is used as the initial condition $I(x, y, 0)$, (x, y) denotes a pixel to be smoothed in the 2-D image domain, t denotes the discrete time steps (iterations). The constant λ is a scalar that determines the stability, and it is usually less than 0.25. $\eta_{(x,y)}$ denotes the spatial neighborhood of pixel (x, y) , and $|\eta_{(x,y)}|$ represents the number of neighbors of point (x, y) . $\nabla I_{(p,q)}^{(x,y)}$

indicates the image intensity difference between two pixels at (x, y) and (p, q) to approximate the image gradient. For the 4-nearest neighbors of point (x, y) as illustrated in Figure 1, the gradients in 4 directions can be calculated in the following way:

$$\begin{aligned}\nabla I_N^{(x,y)} &= I(x, y-1, t) - I(x, y, t) \\ \nabla I_S^{(x,y)} &= I(x, y+1, t) - I(x, y, t) \\ \nabla I_E^{(x,y)} &= I(x+1, y, t) - I(x, y, t) \\ \nabla I_W^{(x,y)} &= I(x-1, y, t) - I(x, y, t).\end{aligned}\quad (6)$$

The choice of the anisotropic diffusion function plays a significant role in preserving image edges. Generally, as proposed in Perona-Malik's work [1], it takes either

$$g(|\nabla I|) = \exp\left[-\left(\frac{|\nabla I|}{k}\right)^2\right] \text{ or } g(|\nabla I|) = 1 / \left(1 + \frac{|\nabla I|^2}{k^2}\right) \quad (7)$$

where $|\nabla I|$ is the gradient magnitude defined in (6), k is a threshold for the gradient magnitude, and it determines the extent to which edges are to be preserved during the anisotropic diffusion process.

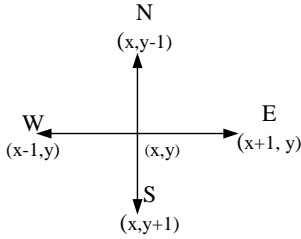


Figure 1. The 4-nearest neighbors of a point in a 2-D grid.

However, when the anisotropic diffusion model is directly applied to the noisy image (it is called the single-scale anisotropic diffusion here), its denoising performance is very susceptible to noise level. From Eq.(6), we can see that when the image $I(x, y, 0)$ contains no or a low-level noise, the high image gradient magnitudes calculated from the difference of image intensity values can surely reflect the existence of edges, and the corresponding small anisotropic diffusion coefficients are reliable. However, when the image is corrupted by a high level of noise, the numerical calculation of the image gradients is very sensitive to noise. In addition to edges, noise may also exhibit high gradients. As a result, for the single-scale anisotropic diffusion, a large gradient magnitude no longer surely corresponds to a true edge and the anisotropic diffusion coefficients are not reliable, resulting in the discounted denoising performance.

So far, much research has been devoted to improving Perona-Malik's anisotropic diffusion method [2-6]. For example, Catte et al. [2] proposed the following anisotropic diffusion equation:

$$\frac{\partial I(x, y, t)}{\partial t} = \text{div}(g(|\nabla(G_\sigma * I)|)\nabla I). \quad (8)$$

The gradients for determining the anisotropic diffusion coefficients are calculated from a "regularized" or smoothed image, which is obtained by filtering the noisy image at each time (iteration) with a Gaussian filter. Similarly, Torkamani-Azar et al. [3] proposed to replace the Gaussian filter with a symmetric exponential filter and the diffusion coefficient is calculated from the convolved image. Although these improvements can convert the ill-posed problem [15] in the Perona-Malik anisotropic diffusion into a well-posed one, their reported denoising performance can be further improved. The problem is that a typical image has a wide variety of edges and it is difficult for one filter to select an optimal scale parameter so as to be adapted to all these edges. For example, a Gaussian filter with a smaller scale parameter can preserve more of the edges, but it cannot

smooth the image sufficiently and noise still has significant influence on the image gradient measurement. On the other hand, using a Gaussian filter with a larger scale parameter for image smoothing, the image can be "regularized" enough, but edges are also smoothed and some weak features may even be removed. As a result, for this kind of regularization method, the calculated gradient magnitudes from the smoothed image may not be able to reflect all true edges in the non-filtered image. So, in both cases, some noisy pixels may be misinterpreted as edge pixels or some true edges may not be detected, and the filtering result may either fail to reduce noise or make the non-detected edges blurred.

Black et al. [4] improved Perona-Malik's anisotropic diffusion from another direction. A different monotonically decreasing function is chosen to determine the anisotropic diffusion coefficient through robust statistics. Compared with the Perona-Malik method [1], the robust anisotropic diffusion demonstrates improved automatic stopping of the diffusion process with preservation of sharp boundaries and better continuity of edges [4]. The new "edge-stopping" function is defined as:

$$g(|\nabla I|, \sigma) = \begin{cases} \frac{1}{2}[1 - (|\nabla I|/\sigma)^2]^2 & |\nabla I| \leq \sigma \\ 0, & \text{otherwise} \end{cases} \quad (9)$$

where $|\nabla I|$ is the gradient magnitude and σ is a threshold about the gradient magnitude. In terms of robust statistics, the "robust scale" σ_e of the image at each time (iteration) is estimated as following [4]:

$$\begin{aligned}\sigma_e &= 1.42826 \text{MAD}(\nabla I) \\ &= 1.42826 \text{median}_I(|\nabla I - \text{median}_I(|\nabla I|)|)\end{aligned}\quad (10)$$

where "MAD" denotes the median absolute deviation. The normalized scale parameter σ in Eq.(9) is defined as $\sigma = \sqrt{5}\sigma_e$ [4]. The constant λ in Eq.(5) for controlling the stability of the anisotropic diffusion is defined as [4]:

$$\lambda = 1/g(\sigma_e, \sigma). \quad (11)$$

Pollak proposed another stabilized inverse diffusion equation (SIDE) to overcome the problem in Perona-Malik's PDE that edges can only be preserved temporarily and will be eventually blurred if the diffusion is allowed for a long time [5]. Wei generalized Perona-Malik's anisotropic diffusion and achieved better denoising performance [6]. However, these algorithms are still a kind of single-scale anisotropic diffusion. When noise level is high, they all have the noise-sensitivity problem.

III. The Proposed Wavelet-based Multiscale Anisotropic Diffusion Algorithm

From the above analysis, we can know it is very necessary for the anisotropic diffusion techniques to reduce the influence of noise on the edge-stopping criterion and gradient measurements. For this end, we propose to decompose the noisy image using the DWT [12-13] so as to construct a linear scale-space representation. After DWT, noise originally in the spatial noisy image is amplified to be the high frequency information, and in the wavelet-based scale-space, noise is mostly located in the finest scale. Furthermore, due to the smoothing functionality of the scaling function in the wavelet transform, noise in the detail subbands tends to decrease as the scale increases. This can be seen from Fig. 2. The wavelet components at each scale are the decomposition results of the approximation component at the next finer scale, while the approximation component at that scale is a smoothed version of the original image. Thus the linear scale-space representation is more stationary than the raw noisy image. Chan et

al developed another wavelet function to represent the piecewise-smooth functions [16], which can be used as an alternative tool for multiscale representation. Afterwards, we perform the MMSE-based filtering on the finest scale. Since for typical images, the real signals are mostly located at the coarser scales and only a small fraction of them, corresponding to the sharpest edges is located at the finest scale, the MMSE-based filtering can significantly reduce noise without affecting edges. As a result, the linear scale-space becomes even more stationary. Finally, we perform the anisotropic diffusion on the more stationary linear scale-space rather than on the raw noisy image domain. Since at each scale, less noise has influence on the PDE than that in the raw noisy image, the anisotropic diffusion coefficients and gradient measurements become more reliable and the anisotropic diffusion is more efficient. Furthermore, the more stationary wavelet-based scale-space makes it possible to optimize the PDE by removing the regularization component such as Gaussian smoothing and other methods [2-3], making the PDE more robust. This means the filtered wavelet-based scale-space works as a new regularization method for the PDE and it is unnecessary to smooth the wavelet-based scale-space using the Gaussian filtering at each step of the anisotropic diffusion process. Gradients can be directly calculated from the wavelet coefficients at the same scale. Zhu et al proposed a multiscale reaction-diffusion method for texture simulation and noise reduction [17], but in their method, the reaction-diffusion is still performed in the spatial domain not in the wavelet transform domain.

A. Dyadic Wavelet Transform based Multiscale Image Analysis

In this work, the translation-invariant two-dimensional (2-D) DWT [12-13] is used to decompose an image I into a linear scale-space. The scale-space consists of an approximation component containing the smoothed structural information of the image at the coarsest scale and a set of detail components describing the details of the image in the horizontal and vertical directions at different scales 2^j , for $j = 1, 2, \dots, J$, with:

$$DWT_J^d I = \left((W_{2^j}^{1,d} I)_{1 \leq j \leq J}, (W_{2^j}^{2,d} I)_{1 \leq j \leq J}, S_{2^j}^d I, J \in Z \right) \quad (12)$$

where d denotes discrete form, $S_{2^j}^d I$ denotes the low-pass component at the coarsest scale 2^j , $W_{2^j}^{1,d} I$ and $W_{2^j}^{2,d} I$ denote the wavelet coefficient $w_{2^j}^{1,d} I(x, y)$ and $w_{2^j}^{2,d} I(x, y)$, respectively, at scale 2^j . In this work, the wavelet function in each orientation is the quadratic spline that approximates the first derivative of a Gaussian function. As a result, the DWT works like the Canny edge detector [18]. Wavelet coefficients with large magnitudes usually correspond to edges in the smoothed image at the next finer scale [12-13], while those with very small magnitudes correspond to the slowly varying regions or constant regions in the image. An example of the 2-D DWT for a piecewise constant image is shown in Figure 2, from which, it is evident that noise in the spatial image domain is mainly located in the finest scale of the scale-space. Also due to the smoothing effect by the low-pass filtering, noise tends to decrease as the scale increases.

B. MMSE-Based Filtering for the Finest Scale

The MMSE-based filtering on the finest scale is to make the wavelet-based scale-space even more stationary.

In the MMSE-based filtering, the image $I(x, y)$ is assumed to be corrupted by the additive white Gaussian noise (AWGN) $n(x, y)$ with variance σ_n^2 . Let us term the observed noisy image $I_n(x, y)$ as:

$$I_n(x, y) = I(x, y) + n(x, y) \text{ for } x, y = 0, 1, 2, \dots, M-1 \quad (13)$$

where M is the image dimension. The DWT [12] of the noisy image $I_n(x, y)$ at scale 2^j , $j \in Z$, can be written as:

$$w_{2^j}^{k,d} I_n(x, y) = w_{2^j}^{k,d} I(x, y) + w_{2^j}^{k,d} n(x, y) \text{ for } x, y = 0, 1, 2, \dots, M-1, k = 1, 2 \quad (14)$$

where $w_{2^j}^{k,d} I(x, y)$ denotes the wavelet coefficient of a noise-free image at location (x, y) and scale 2^j , while $w_{2^j}^{k,d} I_n(x, y)$ denotes the observed wavelet coefficient of the noisy image, and $w_{2^j}^{k,d} n(x, y)$ denotes the wavelet coefficient of the zero-mean and σ_n^2 -variance additive white Gaussian noise.

In this paper, the wavelet coefficients $w_{2^j}^{k,d} I(x, y)$ of the noise-free image at the finest scale ($j = 1$) are assumed to be the conditionally independent zero-mean Gaussian random variables $N(0, \sigma_{(x,y)}^2)$, given their locally varying variances $\sigma_{(x,y)}^2$. This idea is extended from that in the orthogonal wavelet transform domain [19], in which the zero-mean Gaussian distribution is assumed for the orthogonal wavelet coefficients at all scales. The variances $\sigma_{(x,y)}^2$ are modeled as identically distributed, highly correlated random variables. According to the maximum likelihood (ML) estimation, the local variance $\sigma_{(x,y)}^2$ is obtained from the observed local noisy wavelet coefficients [14, 19]:

$$\hat{\sigma}_{(x,y)}^2 = \text{Max}\left\{0, \frac{1}{|\eta_{(x,y)}|} \sum_{(p,q) \in \eta_{(x,y)}} [w_{2^j}^{k,d} I_n(x, y)]^2 - \sigma_n^2\right\} \quad (15)$$

where $\eta_{(x,y)}$ denotes the spatial neighborhood of the position of $w_{2^j}^{k,d} I_n(x, y)$, $|\eta_{(x,y)}|$ denotes the number of neighbors in $\eta_{(x,y)}$. The neighborhood $\eta_{(x,y)}$ is defined as a square window centered at the position of $w_{2^j}^{k,d} I_n(x, y)$. The noise standard deviation σ_n can be estimated directly using a robust estimation, the median absolute deviation of wavelet coefficients in the lowest decomposition level divided by 0.6745 [20],

$$\hat{\sigma}_n = \text{Median}(|w_{2^j}^{k,d} I_n(x, y)|) / 0.6745 \quad (16)$$

After the variance of local noise-free wavelet coefficients is estimated, the noise-free wavelet coefficient value of $w_{2^j}^{k,d} I(x, y)$ is estimated [14, 19]:

$$\hat{w}_{2^j}^{k,d} I(x, y) = \frac{\hat{\sigma}_{(x,y)}^2}{\hat{\sigma}_{(x,y)}^2 + \sigma_n^2} w_{2^j}^{k,d} I_n(x, y). \quad (17)$$

It needs to be pointed out, since the variances are estimated from the local observed noisy wavelet coefficients within a small moving window, the zero-mean assumption is only adequate at the finest scale, and it is inadequate at coarser scales, where when a window contains edges, the mean of wavelet coefficients within the window

is likely non-zero and the estimated $\sigma_{(x,y)}^2$ is not robust. An optimal way is to use a biased-mean model [21].

C. The Rationale of the Wavelet-based Multiscale Anisotropic Diffusion

It is well known that the anisotropic diffusion is best suitable to smooth the piecewise-constant images separated by edges. How can we perform the anisotropic diffusion on the wavelet transform domain? The rationale is that for a typical piecewise-constant image, after it is decomposed using the translation-invariant DWT [12], the wavelet transform components of $(W_{2^j}^{1,d} I)_{1 \leq j \leq J}, (W_{2^j}^{2,d} I)_{1 \leq j \leq J}$ at each scale are still piecewise-constant separated by wavelet coefficients with large magnitudes due to the chosen quadratic spline wavelet function. As shown in Figure 2, wavelet coefficients in the components of $W_2^{1,d} I, W_2^{2,d} I, W_4^{1,d} I,$ and $W_4^{2,d} I$, corresponding to smooth regions in the original image S_1 , are with very small magnitudes, while those corresponding to the vertical and horizontal edges are with very large magnitudes. Thus, the magnitudes of wavelet coefficients reflect the variations in the image intensity values. Therefore, it is possible to perform the anisotropic diffusion on the piecewise-constant wavelet transform components to reduce noise in wavelet coefficients while preserving the edge-related wavelet coefficients.

For the single-scale anisotropic diffusion, the gradients for determining the anisotropic diffusion coefficients are calculated either directly from the raw noisy image [1] or from the Gaussian-smoothed image with the difference of image intensity values [2]. For the multiscale anisotropic diffusion on wavelet transform domain, how can we calculate the gradients for determining the anisotropic diffusion coefficients? We propose to calculate the finite difference (FD) of wavelet coefficients at the same scale as the gradients. If the absolute FD between a central wavelet coefficient and one of its neighboring coefficients is very small, it means that the two wavelet coefficients are located at the same smooth region. Therefore, the anisotropic diffusion coefficient at the corresponding direction will be close to one so that the neighboring wavelet coefficient can be actively involved in smoothing the central wavelet coefficient. On the other hand, if the absolute FD is large, the two wavelet coefficients are separated by an edge and the anisotropic diffusion coefficient will be very small, resulting in the preservation of edge-related wavelet coefficients.

D. Numerical Implementation of Wavelet-based Multiscale Anisotropic Diffusion

In this work, we perform the robust anisotropic diffusion algorithm [4] on the wavelet transform components of $W_{2^j}^{1,d} I$ and $W_{2^j}^{2,d} I$ at different scales forming the wavelet-based multiscale anisotropic diffusion to reduce noise in the wavelet coefficients. Within this framework, any other anisotropic diffusion techniques can be performed on the wavelet transform domain. The proposed anisotropic diffusion process for smoothing the wavelet transform coefficients is defined as:

$$w_{2^j}^{k,d} I(x, y, t+1) = w_{2^j}^{k,d} I(x, y, t) + \frac{\lambda}{|\eta_{(x,y)}|} \sum_{(p,q) \in \eta_{(x,y)}} g(|\nabla w_{2^j}^{k,d} I(x, y, t)|, \sigma) \nabla w_{2^j}^{k,d} I(x, y, t) \quad (18)$$

Compared with the conventional anisotropic diffusion as done in [3], the above equation is more optimal. The Gaussian smoothing component is removed from the PDE due to the stationary DWT representation of the noisy image. Here, $w_{2^j}^{k,d} I(x, y, t)$ denotes the

wavelet coefficient at a position (x, y) at time (iteration) t , $k = 1, 2$, indicating the horizontal and vertical directions, σ is a threshold for the gradient magnitude tuned for a particular application, and $\nabla w_{2^j}^{k,d} I(x, y, t)$ is the gradient, for which the finite differencing scheme is applied to the DWT domain. σ and λ can be automatically calculated using (10) and (11). Consider the neighborhood support illustrated in Figure 1, the gradient $\nabla w_{2^j}^{k,d} I(x, y, t)$ in the 4 directions is calculated in the following way:

$$\begin{aligned} \nabla_N w_{2^j}^{k,d} I(x, y, t+1) &= w_{2^j}^{k,d} I(x, y-1, t) - w_{2^j}^{k,d} I(x, y, t) \\ \nabla_S w_{2^j}^{k,d} I(x, y, t+1) &= w_{2^j}^{k,d} I(x, y+1, t) - w_{2^j}^{k,d} I(x, y, t) \\ \nabla_E w_{2^j}^{k,d} I(x, y, t+1) &= w_{2^j}^{k,d} I(x+1, y, t) - w_{2^j}^{k,d} I(x, y, t) \\ \nabla_W w_{2^j}^{k,d} I(x, y, t+1) &= w_{2^j}^{k,d} I(x-1, y, t) - w_{2^j}^{k,d} I(x, y, t) \end{aligned} \quad (19)$$

The other parameters are calculated in the identical way as done in the robust anisotropic diffusion algorithm [4].

The anisotropic diffusion is iteratively performed on each scale of wavelet transform components $W_{2^j}^{1,d} I$ and $W_{2^j}^{2,d} I$ for noise reduction until a certain number of iterations are reached. Since noise tends to decrease as the scales increase, the iteration number can be set smaller as the scale increases.

E. Summary of the Proposed Algorithm

The proposed algorithm for noise reduction can be summarized as follows:

1. Decompose the noisy image into a scale-space with 4 levels using the DWT to obtain the components $(W_{2^j}^{1,d} I)_{1 \leq j \leq 4}, (W_{2^j}^{2,d} I)_{1 \leq j \leq 4},$ and $S_{2^4}^d I$.
2. For the wavelet transform components of $W_2^{k,d} I$ and $W_2^{2,d} I$, perform the MMSE-based filtering as described in Section III.A. This step can be optional.
3. Perform the anisotropic diffusion algorithm described in Section III.D on the wavelet transform components $(W_{2^j}^{1,d} I)_{1 \leq j \leq 4}$ and $(W_{2^j}^{2,d} I)_{1 \leq j \leq 4}$, to obtain the denoised components $(\hat{W}_{2^j}^{1,d} I)_{1 \leq j \leq 4}$ and $(\hat{W}_{2^j}^{2,d} I)_{1 \leq j \leq 4}$. The components of $(W_{2^j}^{1,d} I)_{j=1}$ and $(W_{2^j}^{2,d} I)_{j=1}$ may have been denoised in Step 2. The low-pass component $S_{2^4}^d I$ is kept without doing any modifications.
4. Perform the inverse DWT on the denoised wavelet transform components, $(\hat{W}_{2^j}^{1,d} I)_{1 \leq j \leq 4}, (\hat{W}_{2^j}^{2,d} I)_{1 \leq j \leq 4}$ and the low-pass component $S_{2^4}^d I$ to reconstruct the denoised image.

The proposed wavelet-based multiscale anisotropic diffusion algorithm with all the 4 steps is called WMSAD and that without including step 2 is called WMSAD-I. The scheme that only includes the MMSE-based filtering on the finest scale is called WT_MMSE.

IV. Experimental Results

The performance of the proposed algorithm is evaluated using the 512x512 standard testing images of Peppers, Lena, Goldhill, and Barbara with 256 gray-scale values. The additive Gaussian white noise with different noise variances is added to these images for performance test. The peak signal to noise ratio (PSNR) values of the 4 noisy images with respect to different noise variances are listed in Table I. Three standard images and their noisy versions with noise

variance 225 are displayed in Figures 3, 4, and 5, respectively. For clarity, only a portion is displayed for each image. For demonstrating the effectiveness of the proposed WMSAD algorithm in noise reduction and edge preservation, the WMSAD algorithm is compared with the counterparts of the robust anisotropic diffusion [4] (RAD), the BSF algorithm [22], and the EWID algorithm [24] in detail from PSNR and the visual quality of the denoised images. The BSF is one of the state-of-the-art wavelet-based denoising techniques while the EWID algorithm is an improved version of the LAWMAP algorithm [19]. When testing the RAD and WMSAD algorithms, the iteration number is set to achieve the maximum PSNR value. The PSNR values of the denoised images for the 4 algorithms with respect to different noise variances are listed in Table II, in which the results of BSF are from the running of the source code [23], and they are higher than those reported in the original paper [22]. From PSNR values, we can see that the proposed WMSAD algorithm achieves the best denoising performance on average among the 4 algorithms. The BSF algorithm achieves a little bit better performance for the image of Barbara, but for all other images, the WMSAD outperforms the BSF. For comparing the visual quality of the denoised images for the 4 algorithms, their denoised images with respect to noise variance 225 are shown in Figures 6, 7, and 8, respectively. By comparing the denoised images shown in Figures 6(a), 7(a), and 8(a), produced by the WMSAD algorithm, with those shown in Figures 6(c), 7(c), and 8(c), produced by the algorithm of RAD, we can see that the WMSAD achieves better visual quality for all images. For RAD, a lot of noise is still remaining in the denoised images. This conforms to our analysis that when the anisotropic diffusion is directly performed on the raw noisy image domain, noise cannot be reduced efficiently due to the influence of noise on the calculation of image gradients. On the other hand, the wavelet-based multiscale anisotropic diffusion is much more immune to noise. Finally by comparing the denoised Lena and Peppers of WMSAD with those of BSF and EWID algorithms, we can see that the WMSAD algorithm achieves much better visual quality. Much less artifacts are on the denoised images. For the denoised Barbara, in order to have a clearer perception for the subtle difference, the eye region in the denoised image is displayed in Figure 9 for WMSAD, BSF and EWID. It can be seen that WMSAD achieves best visual quality. For both BSF and EWID, there are some distortions in the two eyes.

Table I. The PSNR (in dB) of the noisy standard testing images of Lena, Barbara, Peppers and Goldhill with respect to different noise variances.

Image	PSNR (dB) vs. noise variance (σ^2)				
	100	225	400	625	900
Lena	28.20	24.66	22.18	20.27	18.75
Barbara	28.20	24.67	22.19	20.29	18.77
Peppers	28.32	24.81	22.36	20.47	18.96
Goldhill	28.20	24.69	22.22	20.33	18.80

In order to illustrate the impact of MMSE-based filtering at the finest scale on the performance of WMSAD, the schemes of WT_MMSE and WMSAD-I are evaluated. Their PSNR values are listed in Table-II, from which, we can see that the WMSAD-I achieves higher PSNR values than the RAD algorithm. This conforms to our analysis that the wavelet-based scale-space is more stationary than the raw noisy image and the multiscale anisotropic diffusion is more efficient. Also, it is evident that the MMSE-based filtering on the finest scale can reduce a lot of noise, which makes the scale-space even more stationary and greatly helps the WMSAD algorithm outperform the WMSAD-I. Without doing the WT_MMSE,

performing the anisotropic diffusion directly on the finest scale as done in the WMSAD-I is similar to that on the raw noisy image and noise still has significant influence on the gradients.

The proposed WMSAD algorithm is also compared with the recently published work about Gauss curvature-driven diffusion (CCD) [25] and complex steerable wavelets in image denoising (Complex_steer) [26] from the PSNR values. In the CCD algorithm, only the image of Lena is used for experiment. When the PSNR value of the noisy image is 23.54 dB (noise variance < 400 as indicated by Table I), the PSNR value of the denoised image with the CCD algorithm is 27.91 dB, and the PSNR value with the mean curvature evolution method is 25.57 dB. But with the WMSAD algorithm, the PSNR value of the denoised image is 31.40 dB even when the PSNR value of the noisy image is 22.18 dB (noise variance = 400). Also from Table II, it is evident that the WMSAD algorithm is much more efficient than the Complex_steer [26] for the two images. Thus we can say that the WMSAD outperforms the state-of-the-art PDE-based denoising techniques.

Finally, we compare the WMSAD with the algorithm of image denoising using scale mixtures of Gaussians in the wavelet domain (SMG) [27], which is one of the most efficient denoising algorithms published so far. From the PSNR values, for most of the images such as Lena and Barbara, the SMG algorithm achieves much higher PSNR values than the proposed WMSAD algorithm. However for the piecewise-smooth image of Peppers, the WMSAD algorithm achieves much better performance.

As for the computational complexity, the proposed algorithm is more computationally expensive than the conventional anisotropic diffusion techniques. This is due to the fact that for the N -level 2-D DWT, the amount of DWT coefficients is $2N$ times that of the original noisy image. As a result, the anisotropic diffusion at all scales is more time consuming than the conventional anisotropic diffusion techniques. However, the wavelet coefficients at different scales can be processed independently, and if the algorithm is run on a computer with the parallel processing mechanism, the computational complexity will be significantly reduced and it will be less than or comparable to that of the conventional anisotropic diffusion techniques.

V. Conclusion

We have presented a very efficient algorithm that can achieve both efficient noise reduction and edge preservation at the same time due to the following four factors. The first is the construction of the wavelet-based linear scale-space, which is more stationary than the raw noisy image. The second is the MMSE-based filtering performed on the finest scale, making the scale-space even more stationary. The third factor is that the anisotropic diffusion model is optimized by removing the regularization component. The fourth is the anisotropic diffusion is performed on the filtered scale-space rather than on the raw noisy image domain, which makes the anisotropic diffusion coefficients more reliable than those for the single-scale anisotropic diffusion. Experimental results with different noisy images demonstrate that this algorithm outperforms the state-of-the-art anisotropic diffusion techniques and achieves both very high PSNR values and very satisfactory visual quality for the denoised image. It is expected that this algorithm is useful for those denoising applications such as medical imaging where good image visual quality is particularly emphasized. In the future, efforts will be concentrated on reducing the computational complexity of the technique without compromising its denoising performance.

Acknowledgment

The authors are grateful to Lin Zhang for her help in a part of computer programming work in C for implementing the algorithms in the paper.

Table II. Performance (PSNR in dB) of the proposed WMSAD algorithm compared with that of the RAD [4], BSF [22], EWID [24], Complex_Steer [26], SMG [27], WMSAD-I, and WT_MMSE algorithms for the images of Lena, Barbara, Goldhill, and Peppers with respect to different noise variances. The results of RAD and EWID are from the authors' implementations rather than from the original papers. The results of BSF [22] are from the running of the authors' source code which is available in [23].

Scheme	Image	PSNR (dB) vs. noise variance				
		100	225	400	625	900
WMSAD	Lena	34.49	32.69	31.40	30.36	29.47
	Barbara	32.25	30.03	28.43	27.23	26.35
	Peppers	34.11	32.55	31.21	30.05	29.04
	Goldhill	32.53	31.03	29.85	28.87	27.98
BSF [22]	Lena	34.34	32.48	31.16	30.23	29.34
	Barbara	32.52	30.10	28.53	27.35	26.48
	Peppers	33.64	31.99	30.81	29.88	29.09
	Goldhill	32.30	30.45	29.22	28.38	27.72
RAD[4]	Lena	33.38	31.27	28.40	26.75	25.10
	Barbara	31.19	28.63	26.90	25.52	24.80
	Peppers	33.57	31.59	29.76	28.51	27.27
	Goldhill	32.11	30.43	29.01	27.92	26.90
EWID[24]	Lena	34.26	32.35	31.01	29.98	29.16
	Barbara	32.09	29.95	28.51	27.35	26.40
	Peppers	33.51	31.77	30.42	29.34	28.41
	Goldhill	32.48	30.57	29.28	28.32	27.54
WMSAD-I	Lena	34.22	31.73	29.63	27.41	26.42
	Barbara	32.01	29.33	27.57	25.90	24.96
	Peppers	33.90	31.53	29.45	27.28	26.18
	Goldhill	32.32	30.49	28.84	27.47	25.98
WT_MMSE	Lena	32.88	29.74	27.46	25.68	24.24
	Barbara	32.05	29.13	26.96	25.27	23.84
	Peppers	32.70	29.64	27.38	25.59	24.13
	Goldhill	32.28	29.46	27.30	25.57	24.15
Complex_Steer [26]	Lena	32.81	N/A	31.07	N/A	29.69
	Barbara	27.56	N/A	25.51	N/A	24.55
SMG[27]	Lena	35.61	33.90	32.66	31.69	N/A
	Barbara	34.03	31.86	30.32	29.13	N/A
	Peppers	33.77	31.74	30.31	29.21	N/A

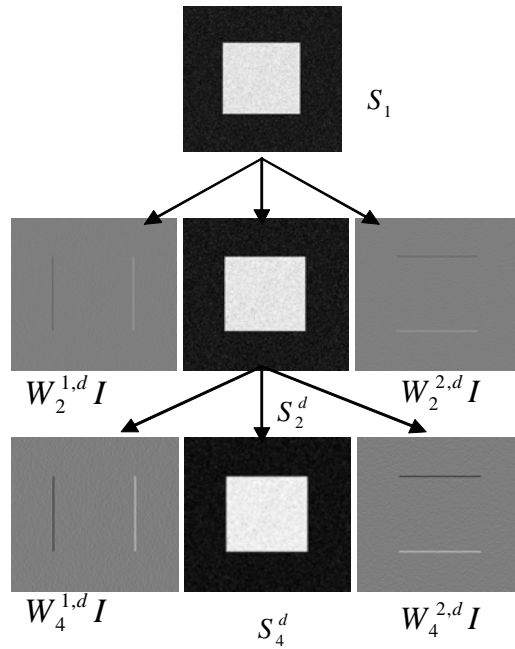


Figure 2. A 2-D DWT decomposition example for a piecewise constant image with 2 decomposition levels. The noisy image is S_1 . $W_2^{1,d}I$ and $W_2^{2,d}I$ denote the DWT coefficients in the first level, while $W_4^{1,d}I$ and $W_4^{2,d}I$ denote the DWT coefficients in the second level. S_2^d and S_4^d are the smoothed versions of the image S_1 .

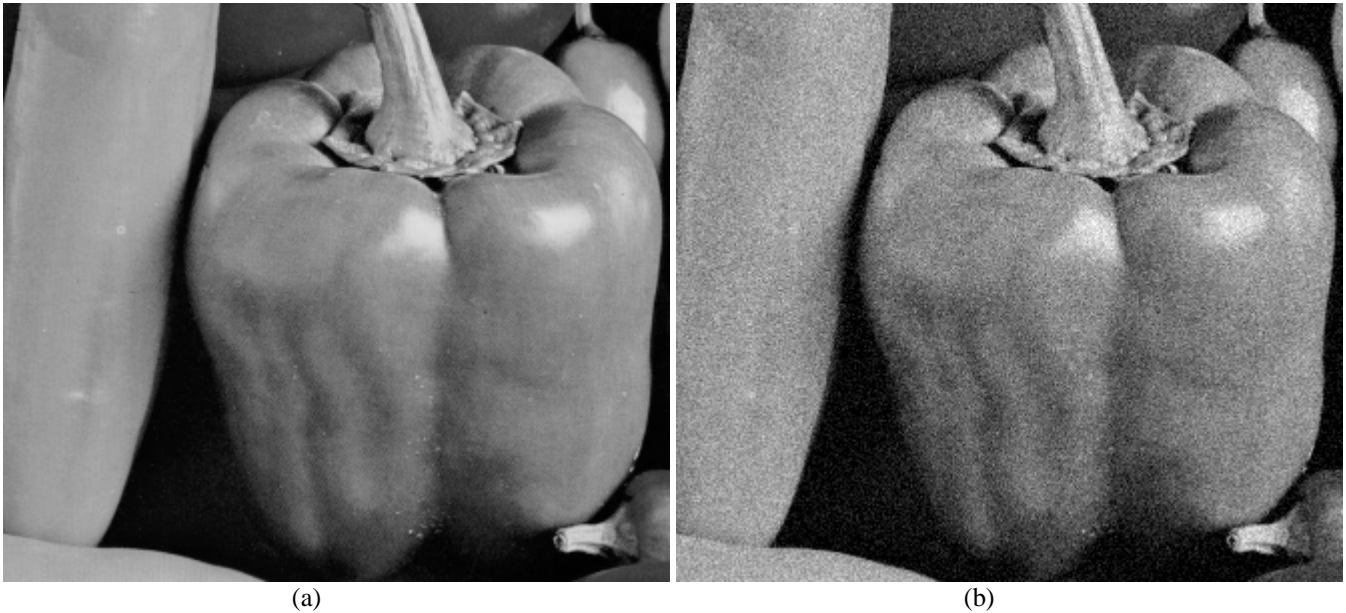


Figure 3. The original and noisy image of Peppers (a portion). (a) is the standard image of Peppers, (b) is the noisy image of Peppers with noise variance of 225.

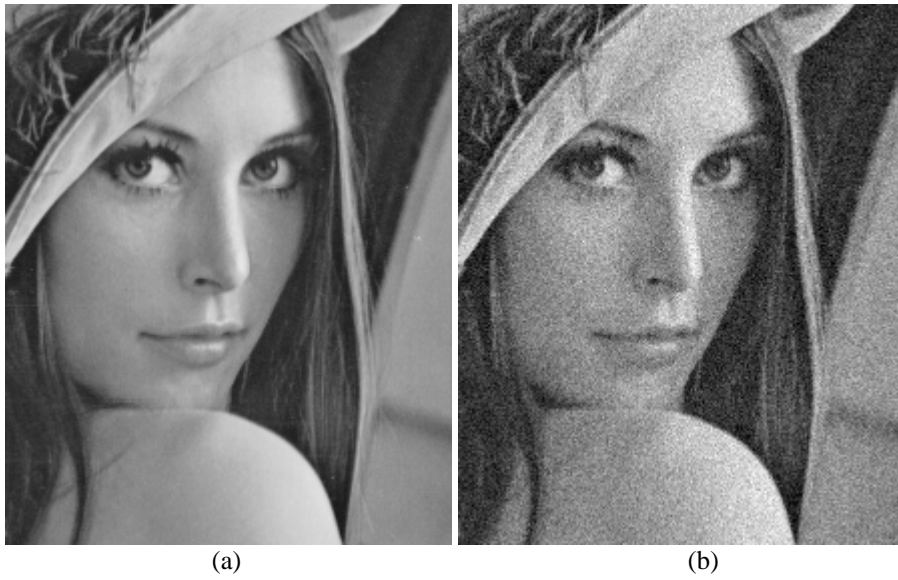


Figure 4. The original and noisy image of Lena (face region). (a) is the standard image, (b) is the noisy image of Lena with noise variance of 225.

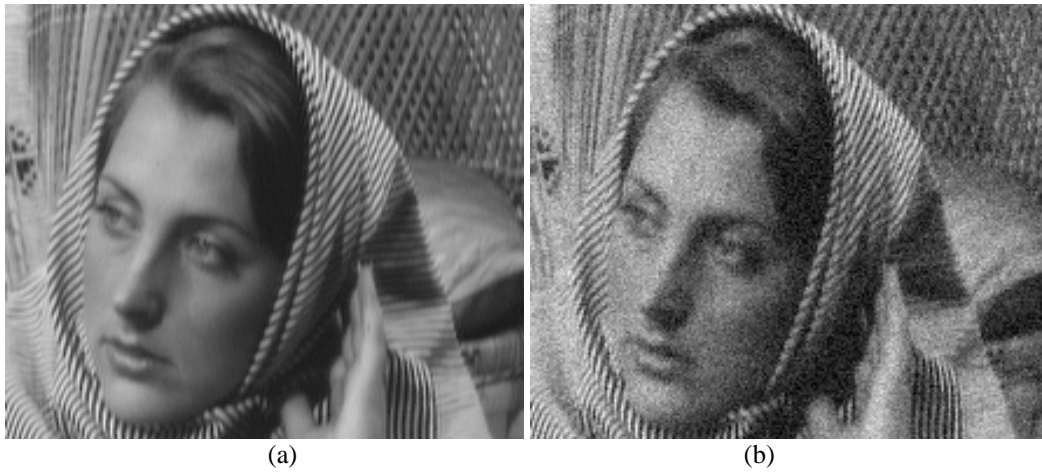


Figure 5. The original and noisy Barbara image. (a) is the standard image of Barbara, (b) is the noisy image of Barbara with noise variance of 225.





Figure 6. The denoised images of peppers using different algorithms for comparison. (a) is with the WMSAD algorithm. (b) is with the BSF [22]. (c) is that of the RAD [4]. (d) is with the EWID[24].

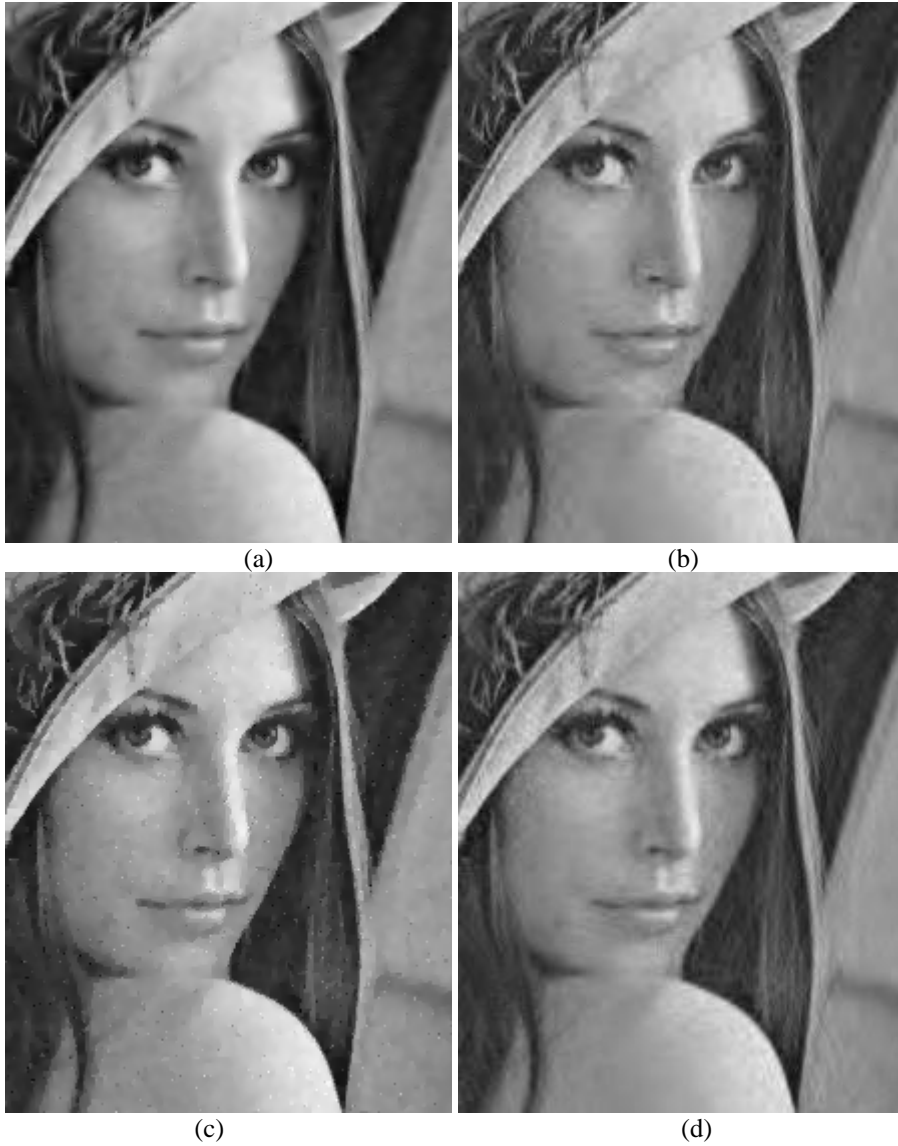


Figure 7. The denoised Lena images (only face region is displayed for clarity) using different algorithms for comparison. (a) is the denoised result with the WMSAD. (b) is that of the BSF[22]. (c) is that of the RAD[4]. (d) is that of the EWID[24].



Figure 8. The denoised images of Barbara (only face region is displayed for clarity). (a) is the denoised result of the WMSAD. (b) is that of the BSF [22]. (c) is that of the RAD[4]. (d) is that of the EWID[24].

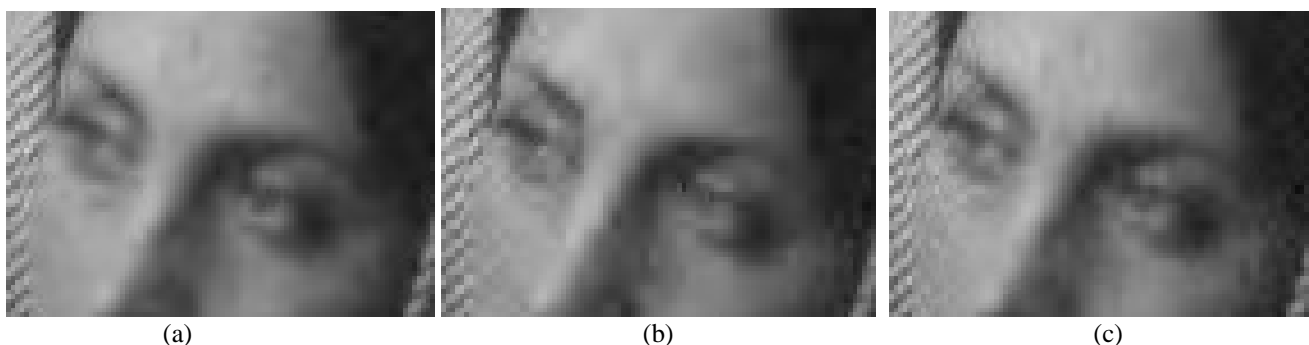


Figure 9. The eye region of the denoised Barbara for different algorithms. (a) is for WMSAD, (b) is for BSF, and (c) is for EWID.

References

- [1] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion", *IEEE Trans. on Pattern Anal. and Mach. Intell.*, vol.12, no. 7, pp.629-639, 1990.
- [2] F. Catte, P. L. Lions, J. M. Morel, and T. Coll, "Image selective smoothing and edge detection by nonlinear diffusion", *SIAM J. Numer. Anal.*, vol. 29, no. 1, pp. 182-193, 1992.
- [3] F. Torkamani-Azar and K. E. Tait, "Image recovery using the anisotropic diffusion equation", *IEEE Trans. on Image Processing*, vol. 5, no. 11, pp. 1573-1578, 1996.
- [4] Michael J. Black, Guillermo Sapiro, David H. Marimont and David Heeger, "Robust anisotropic diffusion", *IEEE Trans. on Image Processing*, vol. 7, no. 3, pp.421-432, 1998.
- [5] I. Pollak, A. S. Willsky, and H. Krim, "Image segmentation and edge enhancement with stabilized inverse diffusion equations", *IEEE Trans. On Image Processing*, vol. 9, no. 2, pp.256-266, 2000.
- [6] Guo W. Wei, "Generalized Perona-Malik Equation for Image Restoration", *IEEE Signal Processing Letters*, vol.6, no.7, pp.165-167, 1999.
- [7] J. Weickert, *Anisotropic diffusion in image processing*, Teubner-Verlag, Stuttgart, Germany, 1998.
- [8] G. Gerig, O. Kubler, R. Kikinis, and F. A. Jolesz, "Nonlinear anisotropic filtering of MRI data", *IEEE Trans. On Medical Imaging*, vol.11, no.2, pp.221-232, 1992.

- [9] M. Lysaker, A. Lundervold, and X-C. Tai, "Noise removal using fourth-order partial differential equation with applications to medical magnetic resonance images in space and time", *IEEE Trans. On Image Processing*, vol.12, no.12, pp. 1579-1590, 2003.
- [10] Z. Ding, J. G. Gore, and A. W. Anderson, "Reduction of noise in diffusion tensor images using anisotropic smoothing", *Magn. Reson. Med.*, vol. 53, pp.585-490, 2005.
- [11] B. Chen, E. W. Hsu, "Noise removal in magnetic resonance diffusion tensor imaging", *Magn. Reson. Med.*, vol. 54, pp.393-407, 2005.
- [12] S. Mallat and S. Zhong, "Characterization of signals from multiscale edges", *IEEE Trans. on Pattern Anal. and Mach. Intell.*, vol.14, no.7, pp.710-732,1992.
- [13] S. Mallat and W. L. Hwang, "Singularity detection and processing with wavelets", *IEEE Trans. on Information Theory*, vol.38, no. 2, pp.617-643, 1992.
- [14] Junmei Zhong, and R. Ning, "Image denoising based on wavelet transform and Multifractals in singularity detection", *IEEE Transactions on Image Processing*, vol. 14, no.10, pp. 1435-1447, Oct. 2005.
- [15] T. H. Romeny B., M., ed., "Geometry-driven diffusion in computer vision", vol. 1 of *Computational Imaging and Vision*, Klumer, 1994.
- [16] T. Chan, H. M. Zhou, "ENO-wavelet transforms for piecewise smooth functions", *SIAM J. on Numerical Analysis*, vol. 40, no.4, pp.1369-1404, 2002.
- [17] S. C. Zhu, D. Mumford, "Prior learning and Gibbs reaction-diffusion", *IEEE Trans. On Pattern Anal. and Mach. Intell.*, vol.19, no.11, pp.1236-1250, 1997.
- [18] J. Canny, "A computational approach to edge detection", *IEEE Trans. Pattern Anal. Mach. Intell.*, vol.8, no. 6, pp.679-697, 1986.
- [19] M. K. Mihcak, I. Kozintsev, K. Ramchandran and P. Moulin, "Low-complexity image denoising based on statistical modeling of wavelet coefficients", *IEEE Signal Processing Letters*, vol. 6, no.12, pp. 300-303, 1999.
- [20] D. L. Donoho and I. M. Johnstone, "Ideal spatial adaptation by wavelet shrinkage", *Biometrika*, vol. 81, no.3, pp.425-455, 1994.
- [21] X. Li and M. T. Orchard, "Spatially adaptive image denoising under overcomplete expansion", *Proceedings of International Conference on Image Processing*, pp.300-303, 2000.
- [22] L. Sendur, I. W. Selesnick, "Bivariate shrinkage functions for wavelet-based denoising exploiting interscale dependency", *IEEE Transactions on Signal Processing*, vol. 50, no. 11, pp. 2744-2756, 2002.
- [23] <http://taco.poly.edu/WaveletSoftware/denoise2.html>.
- [24] Z. Cai, T. H. Cheng, C. Lu and K. R. Subramanian, "Efficient wavelet-based image denoising algorithm", *Electronics Letters*, Vol. 37, No.11, pp.683-685, 2001.
- [25] Suk-Ho Lee and Jin Keun Seo, "Noise removal with Gauss curvature-driven diffusion", *IEEE Trans. on Image Processing*, vol. 14, no. 7, pp. 904-909, 2005.
- [26] A. Anthony Bharath and Jeffrey Ng, "A steerable complex wavelet construction and its application to image denoising", *IEEE Trans. on Image Processing*, vol. 14, no.7, pp. 948-959, 2005.
- [27] J. Portilla, V. Strela, M. J. Wainwright, and E. P. Simoncelli, "Image denoising using scale mixtures of Gaussians in the wavelet domain," *IEEE Trans. Image Process.* vol. **12**, no.11, pp.1338-1351, 2003.

Point-by-point response to reviewers' comments

Dear reviewers,

Thank you very much for the valuable comments. The paper has been revised according to these comments. The point-by-point responses to the comments are given below.

Junmei Zhong & Huifang Sun

Review Number 1.

Comments to the Author

Comment 1-1:

The sole contribution of this paper is the idea of nonlinear diffusion filtering of wavelet coefficients. This contribution is not enough to warrant publication in a Transactions. It is described in Sections III-D and III-E which take up less than a page in the double-column format. The discussion preceding these sections consists mostly of the background material which, for a reader familiar with diffusion filtering and wavelets, does not contribute anything to the understanding of the authors' contribution.

Response 1-1:

The whole Section III is the body of the proposed algorithm. It describes how the idea is formed and the individual components of the proposed algorithm.

Comment 1-2:

The experimental part is very weak. Diffusion-based denoising methods are notoriously sensitive to the number of iterations: if too many iterations are used, the image is oversmoothed resulting in low PSNR; on the other hand, if too few iterations are used, the estimate is noisy, also resulting in a low PSNR. Thus, any diffusion-based denoising algorithm A can be "shown" to be better than another diffusion-based denoising algorithm B by setting the number of iterations for A to achieve the best performance and setting the number of iterations for B carelessly. For example, looking at Fig. 8, it is clear that (a) and (c) were oversmoothed, and (c) has been under-smoothed.

Response 1-2:

For all anisotropic diffusion techniques, it is generally done by setting an iteration number for controlling the diffusion process. This may have the problem as pointed out by the reviewer. How to approach automatic stopping for the anisotropic diffusion is still an open problem and it does need more investigation. In the revised paper, we set the iteration number as many as possible so that the PSNR for the robust anisotropic diffusion is maximized.

Fig. 8(c) looks under-smooth. It is due to the fact that the original anisotropic diffusion is sensitive to noise, and even when the iteration number is increased, some noise still cannot be removed. This conforms to the analysis in this paper that since the model is sensitive to noise, noise can also exhibit high gradients and some noise cannot be smoothed.

Comment 1-3:

Many of these algorithms are also very sensitive to other parameters, such as the time discretization step λ in (19), the function g in (19), and σ in (19). Thus constructing a meaningful experiment to compare several such algorithms is very challenging, and is certainly not achieved in this paper, as these crucial issues of parameter selection are not addressed.

Response 1-3:

The calculation of the parameters in (19) including the scale parameter σ and λ is addressed in detail in the robust anisotropic diffusion [4]. They are all automatically calculated from the image as listed in formula (10) and (11). For the wavelet-based multiscale anisotropic diffusion, the iteration number is set smaller when scale increases.

Comment 1-4:

Finally, diffusion-based denoising has become irrelevant as other, more powerful denoising methods have been developed in the last few years, based on overcomplete representations such as curvelets and on statistical models such as Gaussian scale mixtures.

Response 1-4:

The authors completely agree with the reviewer that denoising techniques based on curvelets and statistical models such as Gaussian scale mixtures are efficient in image denoising. In the mean time, the authors also would like to say that for a wide variety of images, it is still very hard to expect one or two denoising techniques to tackle the challenging problem of noise reduction and edge preservation. For example, noise statistics in MR images, CT images, PET images and ultrasound images are different. For medical image denoising with different modalities, it is impossible for a single denoising technique to efficiently reduce noise for all these image modalities. For biomedical image denoising, when a high PSNR is obtained, image edges must be preserved. When one denoising technique is very efficient for some kinds of images, it may be not very efficient for the other images. For example, the denoising technique based on Gaussian scale mixture is very efficient for many images such as Lena, Barbara, but for the image of Peppers, some other techniques are more efficient than the technique. This is very natural. Recently, there is a lot of research about anisotropic diffusion in biomedical imaging [8-11] for edge-preserving image denoising since the PDE model is theoretically efficient for preserving edges. The motivation of this research is trying to develop an improved anisotropic diffusion technique so that the noise-sensitivity problem as addressed in this paper for the PDE-based anisotropic diffusion can be solved well so that it can be much more useful for emerging biomedical imaging applications.

Review Number 2.

Comments to the Author

The paper proposes an interesting combination of MMSE-based filtering on the finest-scale subbands of a wavelet transform and anisotropic diffusion on the other subbands at coarser scales to denoise images.

Minor Changes:

Comment 2-1:

1/ Pg 2, Eq(7) and Eq(9), can some other symbol be used instead of x , as the latter has previously been used for x -coordinates

Response 2-1:

In (7) and (9), they have been changed into $|\nabla I|$, which denotes the magnitude of gradient as defined in Eq. (6).

Comment 2-2:

2/ Pg 2, line after Eq(7), "where x is the gradient" `_magnitude_`

Response 2-2:

Yes, it has been corrected.

Comment 2-3:

3 / Pg 4, last paragraph, line 8, it may be worth clarifying that as the multiscale wavelet transform has already shifted coarser-scale image information to a coarser spatial-frequency subband, there is no/less of a need to keep applying the Gaussian regularization filter at each step of the diffusion as in [2].

Response 2-3:

Yes, you are right. We have added this. Actually, since the wavelet-based scale-space is more stationary than the raw noisy image, and after the MMSE-based filtering on the finest scale, the scale-space becomes even more stationary. So it is not necessary to do the Gaussian smoothing at each scale. We optimized the PDE by removing the Gaussian smoothing component from the PDE. This is more optimal than in [2]. In the first submission, from the Equation (19), there was no Gaussian smoothing.

Comment 2-4:

4 / Pg 4, next sentence, it may be a good idea to clarify that the finite differences are computed on wavelet coeffs at the same scale.

Response 2-4:

Yes, it has been added in the revised paper.

Comment 2-5:

5 / Pg 5, it would be interesting if the authors could include a graph of the PSNR achieved separately by (a) the MMSE finest-scale filtering and

(b) the anisotropic diffusion on coarser subbands as noise variance is increased. This would give an indication of the contribution of the two components at different noise levels.

Response 2-5:

For (a), we have included the denoising results for only filtering the finest scale using the MMSE-based method, which is called the WT_MMSE in this work. The result is included in Table II.

For (b), the results in Table II for the WMSAD_I algorithm are for the anisotropic diffusion at all scales without doing MMSE-based filtering on the finest scale. Since the direct application of the anisotropic diffusion model to the finest scale is similar to that of the conventional anisotropic diffusion technique, noise still has great influence on the PDE, noise cannot be reduced efficiently. That why the PSNR results of WMSAD-I are lower than those of WMSAD. We hope this makes sense.

Comment 2-6:

6 / Page 7, Table II, As well as complex steer, the authors should include the denoising psnr results for Gaussian Scale Mixture in the Wavelet Domain, by Portilla et al., IEEE Trans Imag Proc, 12(11):133801351, 2003.

Response 2-6.

The algorithm in that paper you mentioned is very efficient for image denoising. The results of that paper are included in the revised paper for comparison.

Comment 2-7:

7/ Page 12, Fig 8, it would be good to have an additional subfigure of the zoomed face region of Lenna as reference.

Response 2-7:

The original image of Lena in Figure 2 is replaced with the face region as reference in the revised paper.

Comment 2-8:

8/ Page 13, Fig 9 (d), why is there a circle on Barbara's "tie"?

Response 2-8: it has been removed.

Review Number 3.

Comments to the Author

A reasonably well written paper, whose main contribution seems to be the application of anisotropic diffusion in a wavelet transform framework. This approach and the supporting results are providing a valid contribution to the field of image denoising and more importantly to the improvement of diffusion based methods. However, a number of improvements or additions should be considered.

Comment 3-1:

Firstly, while the results do offer improvements with respect to robust anisotropic diffusion (and the other referred methods), other perhaps superior wavelet based techniques (in terms of reported PSNR) were not mentioned, (eg. those of Sendur et. al. and Deng et. al.)

Response 3-1:

Sendur et al's paper is cited and compared in detail by running the source code in the website [23] in the revised paper.

Comment 3-2:

This is particularly relevant, given the optional MMSE based filtering proposed by the authors for the finest scale and the significant improvement in PSNR that results. A question that should be addressed is why the proposed anisotropic diffusion approach used at the remaining scales is more suitable than the reported MMSE fine-scale method (used at all scales) or indeed one of the other leading wavelet-based methods.

Response 3-2:

The main reason is that when a noisy image is transformed using the wavelet, noise in the spatial image is amplified to be the high frequency information and is mostly located at the finest scale of the time-frequency DWT domain. If we directly perform the anisotropic diffusion on the finest scale, noise will have great influence on the anisotropic diffusion coefficient and other gradient measurements in the PDE, making the anisotropic diffusion not efficient for noise reduction. On the other hand, in the finest scale, noise is dominant and the real signal components for typical images are very weak especially when noise level in the image is high and they are mostly located at coarser scales. So when we perform the MMSE-based filtering on the finest scale, noise can be reduced significantly while real signal components cannot be affected, making the wavelet-based scale-space even more stationary.

On the other hand, since noise tends to decrease as the scale increases, the coarser scales are more stationary than the raw noisy image. So it is possible to directly perform the anisotropic diffusion on the coarser scales and this can preserve the edge-related wavelet coefficients while reducing noise in wavelet coefficients.

But the MMSE-based filtering cannot be applied to all scales as addressed in the paper. It is due to the fact that in the MMSE-based filtering method, it is assumed that the noise observes the zero-mean Gaussian distribution. For the finest scale, this assumption is adequate, but for coarser scales, this assumption is inadequate since the image structural information is mostly located at the coarser scales. We calculate the parameters using a moving window, at coarser scales, when the window contains edge-related information, the mean of wavelet coefficients within the window may not be zero. As a result, the model is no longer robust at coarser scales. This is confirmed by the LAWMAP algorithm [19] in which the MMSE-based filtering is applied to all scales.

Comment 3-3:

Briefly contrasting the proposed method with leading non-diffusion based approaches would be useful addition. If the proposed method offers improvement in visual quality (or efficiency) rather than PSNR, than this would be a very useful comparison.

Response 3-3:

Sendure et al's paper [22] has been compared in detail with its denoised images by running the code in [23].

Comment 3-4:

The number of images displayed for reader evaluation is perhaps excessive. Key examples demonstrating the desired effects would be adequate. The authors refer to the degree of stationarity of the respective data spaces. This comes across as a rather intuitive assessment rather than something that is being objectively measured. Perhaps this could be clarified by providing a basis for this description (either referenced or explained).

Response 3-4:

Yes, the authors have removed some images by only keeping some key examples as suggested by the reviewer. As for the degree of stationarity in the DWT domain, the authors have re-stated this in the revised paper. Noise is

dominant at the finest scale and tends to decrease as the scale increases. This is due to the smoothing functionality of the scaling function in the wavelet transform. Since the DWT decomposition results at a coarser scale ($j+1$) are from the approximation (low-pass) component S_j^d at the next finer scale (j) as shown in Figure 2, while the approximation component mainly contains the main structure of the image at the corresponding scale (j), and less noise is in S_j^d than in the original noisy image due to the low-pass filtering effect of the scaling function. This property is applicable to all DWT. We hope this makes more sense.