Research Article

Optical Coherence Tomography Noise Reduction Using Anisotropic Local Bivariate Gaussian Mixture Prior in 3D Complex Wavelet Domain

Hossein Rabbani,[1,2] Milan Sonka,[2] and Michael D. Abramoff[2]

[1] *Biomedical Engineering Department, Medical Image & Signal Processing Research Center,*
 Isfahan University of Medical Sciences, Isfahan 81745, Iran
[2] *The Iowa Institute for Biomedical Imaging, The University of Iowa, Iowa City, IA 52242, USA*

Correspondence should be addressed to Hossein Rabbani; h_rabbani@med.mui.ac.ir

In this paper, MMSE estimator is employed for noise-free 3D OCT data recovery in 3D complex wavelet domain. Since the proposed distribution for noise-free data plays a key role in the performance of MMSE estimator, a priori distribution for the pdf of noise-free 3D complex wavelet coefficients is proposed which is able to model the main statistical properties of wavelets. We model the coefficients with a mixture of two bivariate Gaussian pdfs with local parameters which are able to capture the heavy-tailed property and inter- and intrascale dependencies of coefficients. In addition, based on the special structure of OCT images, we use an anisotropic windowing procedure for local parameters estimation that results in visual quality improvement. On this base, several OCT despeckling algorithms are obtained based on using Gaussian/two-sided Rayleigh noise distribution and homomorphic/nonhomomorphic model. In order to evaluate the performance of the proposed algorithm, we use 156 selected ROIs from $650 \times 512 \times 128$ OCT dataset in the presence of wet AMD pathology. Our simulations show that the best MMSE estimator using local bivariate mixture prior is for the nonhomomorphic model in the presence of Gaussian noise which results in an improvement of 7.8 ± 1.7 in CNR.

1. Introduction

Optical coherence tomography (OCT) is an optical signal acquisition and processing method that captures 3D images from within optical scattering media such as biological tissues [1–4]. For example, in ophthalmology, OCT is used to obtain detailed images from within the retina [4]. Similar to other optical tomographic techniques, OCT suffers from speckle noise that reduces the ability of image interpretation [5]. So, noise reduction is an essential part of OCT image processing systems. Until now, several techniques for OCT noise reduction have been reported [6–14]. Initial methods perform in complex domain [15], that is, before producing magnitude of OCT interference signal, while most introduced despeckling methods are applied after an OCT image is formed [6–14]. These methods, which usually suppose multiplicative noise

for speckled data, also can be categorized into image domain and transform domain methods. As an example for image domain techniques, in [16] the rotating kernel transform (RKT) filters are applied on an image with a set of oriented kernels and keep the largest filter output for each pixel. Other image domain methods based on enhanced Lee filter [17], median filter [17], symmetric nearest neighbor filter [17], and adaptive Wiener filter [17], and I-divergence regularization [6] and PDE-based nonlinear diffusion methods [14] have been reported in the literature.

Transform domain techniques typically outperform the image domain techniques because incorporating speckle statistics in the despeckling process would be facilitated in sparse domains. Such techniques apply a sparse transform (such as wavelet and curvelet transforms) [7–12, 18] directly on data (viz., nonhomomorphic methods) or on

log-transformed data (viz., homomorphic methods), and suppose that in the sparse domain noise is converted to additive white Gaussian noise (AWGN) [13] or other models which can be removed using an appropriate shrinkage function. For example, in [18], a spatially adaptive wavelet thresholding method is used for speckle suppression in log-transformed domain. Since actual signal in OCT images consists of horizontal edges arising from reflections at the layer boundaries, most of the edge information is in "low-pass"-"high-pass" (LH) subbands (and some of it in HH subbands). Therefore, an increased threshold in the vertical subbands using a constant multiplier ($K = 4$) is chosen to decrease further noise with a minimal effect on the edge sharpness. Other transform domain methods based on hard thresholding in 3D curvelet domain [8], soft thresholding in discrete complex wavelet transform (DCWT) domain [9], and temporal and spatial wavelet-based filtering [10] have been reported in other literatures.

In fact denoising is the problem of obtaining the noise-free data from noisy data observation, which may be solved in a deterministic or probabilistic framework. In the first case, each voxel is considered as an unknown deterministic variable, and non-Bayesian techniques are employed to solve this problem. In the second case, the data is modeled as a random field, and Bayesian methods are used for the estimation of clean data from the noisy environment. Therefore, the proposed prior probability distributions for noise-free data and noise (i.e., proposed as speckle for OCT data) play a key role in the noise reduction problem.

1.1. Statistical Properties of Noise-Free Coefficients.

Description of the statistical properties of natural signals can be facilitated in the wavelet domain [19] due to sparseness and decorrelation properties of wavelets [20]. The sparseness property states that the marginal pdf of wavelet coefficients in each subband has a large peak at zero and its tails fall to zero slower than the Gaussian pdf (leptokurtic). On this base, some long-tailed pdfs such as generalized Gaussian distribution (GGD) [21, 22], α-stable distributions [23], Bessel K form densities [24, 25], and mixture pdfs [26–31] have been proposed. Although the decorrelation property of wavelets states that coefficients at the same positions in the adjacent scales are uncorrelated, it does not mean that they are independent. The interscale dependency of wavelet states that large/small values of wavelet coefficients tend to propagate across scales [32]. Some researchers have proposed hidden Markov models (HMMs) [33] and Markov random fields (MRFs) [34] to model the interscale dependency [35]. Recently, it has been shown that some non-Gaussian bivariate joint pdfs for each coefficient and its parent, such as circular symmetric Laplacian pdf [36], bivariate Cauchy distribution [37], (multivariate) Gaussian scale mixture (GSM) model [27, 38, 39], and bivariate Laplacian mixture models [40] are able to capture this property easily and produce better denoising results with lower computational complexity.

The dependencies between wavelet coefficients are not restricted to the interscale dependency. There is another dependency between spatial adjacent coefficients in each subband, namely, intrascale dependency [42]. This dependency states that if a particular wavelet coefficient is large/small, then the spatial adjacent coefficients are likely to be large/small too. Usually this property is captured using local parameters for pdfs [37], and it has been shown that denoising algorithms using this property for statistical modeling of wavelets are able to improve the denoising results [43–45]. For example, Mihçak [43] employs local variance for Gaussian pdf to model intrascale dependency. In [44], a mixture of two Laplace pdf with local parameters is proposed for simultaneously capturing heavy-tailed nature and intrascale dependency. Reference [45], using local variance for proposed model in [36], improves the results for noise reduction application because this local pdf models both interscale and intrascale dependencies. In this paper, we extend the proposed pdf in [46] based on a mixture of bivariate Gaussian pdfs with local parameters for noise-free wavelet coefficients. Since the empirically observed distribution of wavelet coefficient pairs in adjacent scales have elliptical symmetry, we use different variances for marginal pdfs that lead to an elliptical symmetric bivariate pdf instead of circular symmetric pdf. Recently, it has been shown that using anisotropic window instead of square window can improve the denoising results [47]. Based on the special structure of OCT data, we choose an anisotropic windowing procedure for local parameters estimation that results in visual quality improvement.

1.2. Discrete Complex Wavelet Transform (DCWT).

The wavelet based image denoising consists of the following steps.

(1) Signal transformation of the noisy observation.

(2) Modification of the noisy wavelet coefficients based on some criteria.

(3) Inverse signal transformation of modified coefficients.

As explained earlier, the second step depends on the type of estimator and for a minimum mean square error (MMSE) estimator, the proposed model for signal and noise (which we propose as a multiplicative model), the proposed pdf of noise-free wavelet coefficients (modeled, in this paper, as a mixture of bivariate Gaussian pdfs with local parameters), and the proposed pdf for noise (with which we test both Gaussian and two-sided Rayleigh distributions) define the performance of the algorithm. However, for the first and last steps of wavelet-based denoising algorithm, the type of transformation plays a key role. In this paper, we use DCWT [48] instead of ordinary discrete wavelet transform (DWT). Despite DWT being a sparse representation that outperforms many signal processing approaches, it does not lead to an optimum performance in all applications and suffers from several fundamental shortcomings (especially in high-dimensional cases), which DCWT avoids them. These shortcomings are as follows.

(1) In the neighborhood of an edge, the DWT produces both large and small wavelet coefficients. In contrast, the magnitudes of DCWT coefficients are

more directly related to their adjacency to the edge. The main reason of this phenomenon is using band-pass filters that produce DWT coefficients which oscillate positively and negatively around the singularities, and this subject complicates wavelet-based processing.

(2) DWT is not shift invariant. It means that a small shift in the input signal of DWT makes the total energy of wavelet coefficients in subband completely differ. This shift greatly perturbs oscillation pattern around singularities of the DWT coefficient which complicates wavelet-domain processing.

(3) Since the DWT coefficients in each subband are produced via critical sampling after using nonideal low-pass and high-pass filters, substantial aliasing would be produced. If the wavelet coefficients are not changed, the inverse DWT cancels this aliasing. Applying any processing method on wavelet coefficients (such as thresholding) disarranges this balance between the forward and inverse transforms which leads to artifacts in the reconstructed signal.

(4) The directional selectivity of 2D DCWT has been explained in Appendix A. Similar to the 2D case, the standard 3D data transforms, which are separable multiplication of 1D tensors, do not provide useful representations with good energy compaction property for 3D data. For example, the multi-dimensional standard separable DWT mixes orientations and motions in its subbands and produces the checkerboard artifacts (Figure 1). In contrast, since the spectrum of the (approximately) analytic 1D wavelet is supported on only one side of the frequency axis, the spectrum of the DCWT in 3D domain is supported in only 1/27 of the 3D frequency plane. So, instead of 3D DWT, usually oriented transforms such as 3D DCWT are proposed for 3D data processing [41, 48–52]. Figure 1 shows a comparison between subbands of 3D DWT and 3D DCWT.

1.3. Organization of the Paper. In Section 2, we explain our proposed pdf for noise-free 3D DCWT coefficients, that is, a mixture of bivariate Gaussian distributions with local parameters. In Section 3, at first we obtain a local thresholding function supposing a priori distribution as a bivariate Gaussian pdf with local variance, and then in a Bayesian framework we produce our new shrinkage functions derived from the proposed pdf and using Gaussian/two-sided Rayleigh noise distribution and homomorphic/non-homomorphic model. In Section 4, we explain the proposed anisotropic window selection procedure for local parameter estimation based on special structure of OCT data. In Section 5, we use our model for wavelet-based denoising of several 3D OCT data. We compare our methods visually and in terms of PSNR. Also in this section, we use the proposed method for nonspeckle noise reduction. Finally, in Section 6, we summarize this paper and suggest some future work.

2. Bivariate Gaussian Mixture Model with Local Parameters

One of the primary properties of the wavelet transform is compression. This property means that the marginal distributions of wavelet coefficients are highly kurtotic, and so long-tailed distributions are suitable models for marginal pdf. A zero-mean mixture model could have a large peak at zero and would be long tailed. For example, in [22, 26, 29, 31] a mixture of Gaussian distributions is proposed to model the heavy-tailed nature of wavelet coefficients. Figure 2 shows this model that consists of two zero-mean Gaussian distributions with two different variances. The Gaussian pdf with low variance can model the large peak at zero and the Gaussian pdf with high variance can model tails of distribution. The secondary properties of the wavelet transform are clustering and persistence. The clustering property, that is called intrascale dependency, states that if a particular wavelet coefficient is large/small, then adjacent coefficients are very likely to also be large/small [36], and usually local pdfs are able to model this property. The persistence property, that is called the interscale dependency, states that large/small values of wavelet coefficients tend to propagate across scales [36]. As an example, Figure 3 illustrates the empirical joint parent-child histogram of wavelet coefficients computed from the 200, 512×512 images from the Corel image database [42]. Usually this property can be modeled using proper bivariate pdfs.

2.1. Description of the Proposed Model. In this paper, we assume a pdf as a mixture of two bivariate Gaussian pdfs with local parameters in order to model the distribution of wavelet coefficients of images as follows:

$$
\begin{aligned}
p_{\overline{w}(k)}\left(\overline{w}\left(k\right)\right) &= a\left(k\right) p_1\left(\overline{w}\left(k\right)\right) + \left(1 - a\left(k\right)\right) p_2\left(\overline{w}\left(k\right)\right) \\
&= \frac{a\left(k\right) e^{\left(w_1^2(k)/2\sigma_{11}^2(k)\right) - \left(w_2^2(k)/2\sigma_{12}^2(k)\right)}}{2\pi\sigma_{11}\left(k\right)\sigma_{12}\left(k\right)} \\
&\quad + \frac{\left(1 - a\left(k\right)\right) e^{\left(-w_1^2(k)/2\sigma_{21}^2(k)\right) - \left(w_2^2(k)/2\sigma_{22}^2(k)\right)}}{2\pi\sigma_{21}\left(k\right)\sigma_{22}\left(k\right)},
\end{aligned}
\tag{1}
$$

where $a(k) \in [0,1]$, $\sigma_{11}(k)$, $\sigma_{12}(k)$, $\sigma_{21}(k)$, $\sigma_{22}(k)$ are the mixture model parameters. For each random bivariable, the second component is the parent of the first component; for example, $w_2(k)$ represent, the parent of $w_1(k)$ at the same spatial position as the kth wavelet coefficient $w_1(k)$ and at the next coarser scale.

Our proposed model in this paper, that is a mixture of bivariate Gaussian pdfs with local parameters, is mixture, bivariate and local. Therefore, it is able to simultaneously capture the heavy-tailed property and inter- and intrascale dependencies.
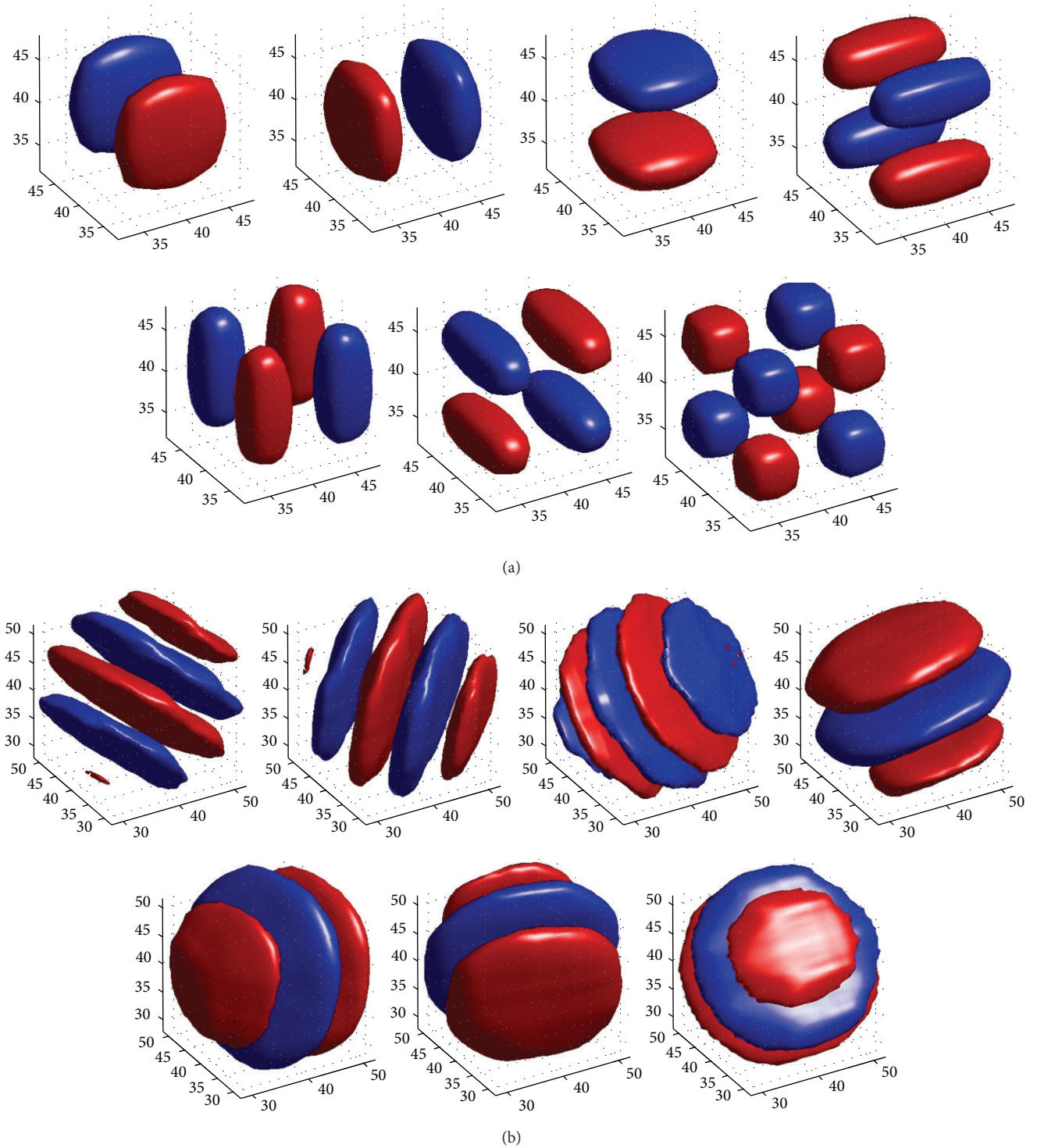
(a)

(b)

FIGURE 1: A comparison between the idealized support of the Fourier spectrum of each standard and complex wavelet in the 3D frequency domain. (a) Isosurfaces of the 7 3D wavelets for a standard 3D wavelet transform. The blue and red colors have the same amplitude, but their phases are complement. (b) Isosurfaces of 7 of the 28 3D wavelets for a 3D DCWT. Each subband corresponds to motion in a specific direction [41].
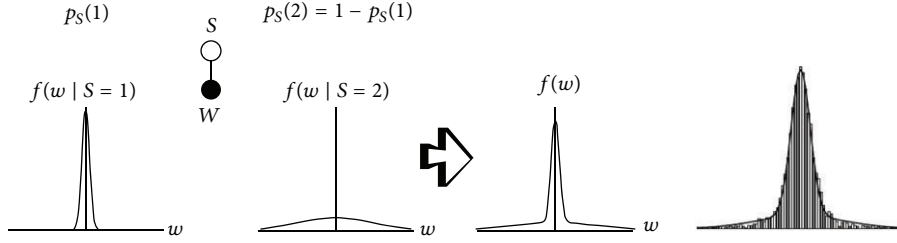
FIGURE 2: Zero-mean Gaussian mixture model (left image) and empirical histogram of wavelet in a subband together with the Gaussian mixture model (right image) [26].
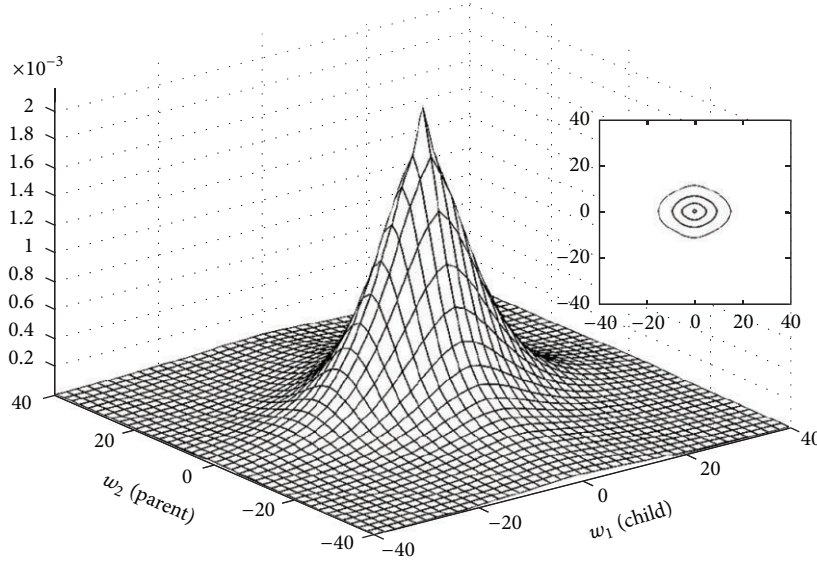

FIGURE 3: Empirical joint parent-child histogram of wavelet coefficients (computed from the Corel image database) [42].

After substitution of mixture model in the definition of $E(w_1(k)w_2(k))$, we can see that this pdf represents two uncorrelated random variables as follows:

$$
\begin{aligned}
E\left(w_1(k) w_2(k)\right) \\
&= \iint w_1(k) w_2(k) p_{\overline{w}(k)}(\overline{w}(k)) \, d\overline{w}(k) \\
&= (1 - a(k)) \\
&\quad \times \iint w_1(k) w_2(k) p_2(\overline{w}(k)) \, dw_1(k) \, dw_2(k) \\
&\quad + a(k) \iint w_1(k) w_2(k) p_1(\overline{w}(k)) \, dw_1(k) \, dw_2(k) \\
&= 0.
\end{aligned}
\tag{2}
$$

Interestingly, the marginal pdf of $w_i(k)$ for $i = 1, 2$ is the mixture of two univariate Gaussian pdf with local parameters [44],

$$
\begin{aligned}
p_{w_1(k)}\left(w_1(k)\right) &= \int_{-\infty}^{\infty} p_{\overline{w}(k)}(\overline{w}(k)) \, dw_2(k) \\
&= a(k) \frac{\exp\left(-w_1^2(k) / 2\sigma_{11}^2(k)\right)}{\sigma_{11}(k) \sqrt{2\pi}}
\end{aligned}
$$

$$
+ (1 - a(k)) \frac{\exp\left(-w_1^2(k) / 2\sigma_{21}^2(k)\right)}{\sigma_{21}(k) \sqrt{2\pi}}.
\tag{3}
$$

It is easy to see that $w_1(k)$, $w_2(k)$ are not independent; that is,

$$
p_{\overline{w}(k)}(\overline{w}(k)) \neq p_{w_1(k)}\left(w_1(k)\right) p_{w_2(k)}\left(w_2(k)\right).
\tag{4}
$$

See Appendix B for more explanation.

2.2. Local EM Algorithm. To characterize the parameters in (1), it is necessary to have the parameters $\sigma_{11}(k)$, $\sigma_{21}(k)$, $\sigma_{12}(k)$, $\sigma_{22}(k)$, and $a(k)$. For this mixture model, we use an iterative numerical algorithm to estimate these parameters. The expectation maximization (EM) algorithm is most frequently used to estimate such parameters. Usually, the EM algorithm for mixture models employs all data in each subband to obtain the parameters. Using this global EM algorithm, equal parameters are obtained for all data in each subband. However, to model the intrascale dependency, we must incorporate the local statistics and need to have different parameters for each voxel in each subband. So, we introduce a local version of EM algorithm. This local EM

algorithm is able to obtain separate parameters for each voxel by the implementation of EM algorithm in each window $N(k)$ centered at $w(k)$. This iterative algorithm has two steps. Assuming that the observed data $\overline{w}(k)$ for $k = 1, \ldots, N$, the E-step calculates the responsibility factors for each data as follows:

$$r_1(k) \longleftarrow \frac{a(k) p_1(\overline{w}(k))}{a(k) p_1(\overline{w}(k)) + (1 - a(k)) p_2(\overline{w}(k))},$$

$$r_2(k) \longleftarrow 1 - r_1(k). \tag{5}$$

The M-step updates the parameters $a(k)$, $\sigma_{11}(k)$, $\sigma_{12}(k)$, $\sigma_{21}(k)$, and $\sigma_{22}(k)$. $a(k)$ is computed by

$$a(k) \longleftarrow \frac{1}{M} \sum_{j \in N(K)} r_1(j), \tag{6}$$

where M is the number of coefficients in the square window $N(k)$ centered at $w(k)$.

The variances $\sigma_{11}(k)$, $\sigma_{12}(k)$, $\sigma_{21}(k)$, and $\sigma_{22}(k)$ are computed by [40]:

$$\sigma_{im}^2(k) \longleftarrow \frac{\sum_{j \in N(K)} r_i(j) w_m^2(k)}{\sum_{j \in N(K)} r_i(j)}, \quad i, m = 1, 2. \tag{7}$$

3. Denoising Using MMSE Estimator

In this section, the denoising of a 3D OCT data is considered. We assume that dominant noise in OCT data is speckle. In this case as a common model, we propose multiplicative model as follows:

$$x(i) = s(i) g(i), \tag{8}$$

where i is the index of voxel and is between 1 and number of voxels.

As explained in Introduction, reported transform-based OCT noise reduction methods in the literatures [7–12, 18] usually at first transform data into log domain, and suppose that noise in log domain is AWGN:

$$W(\log x(i)) = W(\log s(i)) + W(\log g(i)), \tag{9}$$

where in this paper W shows 3D DCWT operator. So, we can write

$$y(k) = w(k) + n(k), \tag{10}$$

where $w(k)$, $y(k)$, and $n(k)$ are, respectively, the kth noise-free 3D DCWT coefficients, noisy 3D DCWT coefficients, and noise in the 3D DCWT domain.

Recently, it has been reported [56–59] that non-homomorphic techniques that do not use this nonlinear operation and apply wavelet transform directly on speckled data lead to unbiased estimation of the data and decrease the computational complexity. On this base after applying 3D DCWT (directly) on data, we would have

$$W(x(i)) = W(s(i) g(i)) = W(s(i) + s(i)(g(i) - 1))$$

$$= W(s(i)) + W(s(i)(g(i) - 1)). \tag{11}$$

Again we can write

$$y(k) = w(k) + n(k), \tag{12}$$

where $w(k)$, $y(k)$, and $n(k)$ are, respectively, the kth noise-free 3D DCWT coefficients, noisy 3D DCWT coefficients, and noise in the 3D DCWT domain. Since speckle noise g can be modeled as a unit-mean random process independent of the noise-free data, we would have $E[W(s(g - 1))] = 0$, and also it can be easily shown [58] that $E[W(s)W(s(g - 1))] = 0$ which means that $w(k)$ and $n(k)$ are zero-mean uncorrelated random variables. If $w_p(k)$, $y_p(k)$, and $n_p(k)$ show the parent coefficients of $w(k)$, $y(k)$, and $n(k)$, respectively, we can write

$$y_p(k) = w_p(k) + n_p(k). \tag{13}$$

Based on the persistence property, we need to have a bivariate model based on parent-child pairs. So, we can propose the following bivariate model:

$$\overline{y}(k) = \overline{w}(k) + \overline{n}(k), \tag{14}$$

where $\overline{w}(k) = (w(k), w_p(k))$, $\overline{y}(k) = (y(k), y_p(k))$, and $\overline{n}(k) = (n(k), n_p(k))$ are, respectively, the kth parent-child pairs of noise-free 3D DCWT coefficients, noisy 3D DCWT coefficients, and additive noise in the 3D DCWT domain. In the literature, several models such as K-distribution, Rayleigh, Weibull, log-normal, and Nakagami distributions have been proposed [57, 58, 60–63] for speckle in image domain. In this paper, we test both AWGN and two-sided Rayleigh model for noise in wavelet domain as follows:

$$p_{\overline{n}}(\overline{n}(k)) = \frac{1}{2\pi\sigma_n^2} \exp\left(-\frac{n_1^2(k) + n_2^2(k)}{2\sigma_n^2}\right), \tag{15}$$

$$p_{\overline{n}}(\overline{n}(k)) = \frac{|n_1(k) n_2(k)|}{4\alpha^4} \exp\left(-\frac{n_1^2(k) + n_2^2(k)}{2\alpha^2}\right), \tag{16}$$

where $\sigma_n^2 = 2\alpha^2$ shows the noise variance.

Now our goal is the estimation of $\overline{w}(k)$ from $\overline{y}(k) = \overline{w}(k) + \overline{n}(k)$, where $\overline{n}(k)$ is a Gaussian or two-sided Rayleigh according to some criteria.

If we employ the MMSE estimator for the estimation problem, we get the posterior mean as an optimal solution:

$$\widehat{w}(k) = \iint w(k) p_{\overline{w}(k)|\overline{y}(k)}(\overline{w}(k) \mid \overline{y}(k)) d\overline{w}(k)$$

$$= \iint w(k) \frac{p_{\overline{y}(k)|\overline{w}(k)}(\overline{y}(k) \mid \overline{w}(k)) p_{\overline{w}(k)}(\overline{w}(k))}{p_{\overline{y}(k)}(\overline{y}(k))} d\overline{w}(k)$$

$$= \frac{\iint w(k) p_{\overline{y}(k)|\overline{w}(k)}(\overline{y}(k) \mid \overline{w}(k)) p_{\overline{w}(k)}(\overline{w}(k)) d\overline{w}(k)}{p_{\overline{y}(k)}(\overline{y}(k))}$$

$$= \frac{\iint w(k) p_{\overline{y}(k)|\overline{w}(k)}(\overline{y}(k) \mid \overline{w}(k)) p_{\overline{w}(k)}(\overline{w}(k)) d\overline{w}(k)}{\iint p_{\overline{y}(k)|\overline{w}(k)}(\overline{y}(k) \mid \overline{w}(k)) p_{\overline{w}(k)}(\overline{w}(k)) d\overline{w}(k)}$$

$$= \frac{\iint w(k) p_{\overline{n}}(\overline{y}(k) - \overline{w}(k)) p_{\overline{w}(k)}(\overline{w}(k)) d\overline{w}(k)}{\iint p_{\overline{n}}(\overline{y}(k) - \overline{w}(k)) p_{\overline{w}(k)}(\overline{w}(k)) d\overline{w}(k)}. \tag{17}$$

3.1. Denoising Based on Modeling Noise-Free Data by Bivariate Gaussian PDF with Local Variance. In order to solve (17), we must know the prior distribution of 3D DCWT coefficients, that is, $p_{\overline{w}(k)}(\overline{w}(k))$. Defining $\mathrm{Gauss}(x,\sigma) :=$ $\exp(-x^2/(2\sigma^2))/(\sigma\sqrt{2\pi})$, if we suppose that $w(k)$, $w_p(k)$ are independent Gaussian pdf with variances $\sigma_1(k)$ and $\sigma_2(k)$, the following bivariate Gaussian pdf with local variances can be proposed for the noise-free wavelet coefficients:

$$p_{\overline{w}(k)}(\overline{w}(k)) = p_{w(k)}(w(k)) \cdot p_{w_p(k)}(w_p(k))$$

$$= \mathrm{Gauss}(w(k),\sigma(k)) \cdot \mathrm{Gauss}(w_p(k),\sigma_p(k))$$

$$\Longrightarrow p_{\overline{w}(k)}(\overline{w}(k))$$

$$= \exp\left[-\frac{w^2(k)}{2\sigma^2(k)} - \frac{w_p^2(k)}{2\sigma_p^2(k)}\right]$$

$$\times \left(2\pi\sigma(k)\,\sigma_p(k)\right)^{-1}. \tag{18}$$

In this case, $w(k)$ and $w_p(k)$ are uncorrelated and independent, and therefore the MMSE estimator of $w(k)$, $w_p(k)$ yields the shrinkage function corresponding to univariate Gaussian pdf, that is, Wiener filter [21] as follows:

$$\widehat{w}(k) = \frac{\iint w(k)\, p_{\overline{n}}(\overline{y}(k) - \overline{w}(k))\, p_{\overline{w}(k)}(\overline{w}(k))\, d\overline{w}(k)}{\iint p_{\overline{n}}(\overline{y}(k) - \overline{w}(k))\, p_{\overline{w}(k)}(\overline{w}(k))\, d\overline{w}(k)}$$

$$= y(k)\, \frac{\sigma^2(k)}{\sigma^2(k) + \sigma_n^2}. \tag{19}$$

And so we can write

$$\widehat{\overline{w}}(k) = \left(\frac{y(k)\,\sigma^2(k)}{\sigma^2(k) + \sigma_n^2},\ \frac{y_p(k)\,\sigma_p^2(k)}{\sigma_p^2(k) + \sigma_n^2}\right). \tag{20}$$

Similarly, if we choose two-sided Rayleigh pdf for noise distribution, the following estimator is obtained [44]:

$$\widehat{w}(k) = 2z(k)\,\sqrt{2}\left(2 - \frac{\sigma^2(k)}{\alpha^2}\right) + \sqrt{\frac{\pi}{2}}\left(1 - 2\frac{\sigma^2(k)\,z^2(k)}{\alpha^2}\right)$$

$$\times (\mathrm{erfc}\,x(z(k)) - \mathrm{erfc}\,x(-z(k)))$$

$$\times \left(\sqrt{\frac{1}{\alpha^2} + \frac{1}{\sigma^2(k)}}\,(2 + z(k)\,\sqrt{\pi}\,\mathrm{erfc}\,x(-z(k))\right.$$

$$\left. - z(k)\,\sqrt{\pi}\,\mathrm{erfc}\,x(z(k))\right)^{-1}, \tag{21}$$

where

$$z(k) = \frac{y(k)}{\sigma^2(k)}\sqrt{\frac{1}{2/\alpha^2 + 2/\sigma^2(k)}}, \tag{22}$$

$$\mathrm{erfc}\,x(u) = \frac{2}{\sqrt{\pi}}\int_0^\infty e^{-t^2 - 2tu}\,dt.$$

And so we can write

$$\widehat{\overline{w}}(k) = \left(\left(2z(k)\,\sqrt{2}\left(2 - \frac{\sigma^2(k)}{\alpha^2}\right)\right.\right.$$

$$+ \sqrt{\frac{\pi}{2}}\left(1 - \frac{\sigma^2(k)\,z^2(k)}{\alpha^2}\right)$$

$$\left.\times (\mathrm{erfc}\,x(z(k)) - \mathrm{erfc}\,x(-z(k)))\right)$$

$$\times \left(\sqrt{\frac{1}{\alpha^2} + \frac{1}{\sigma^2(k)}}\right.$$

$$\times (2 + z(k)\,\sqrt{\pi}\,\mathrm{erfc}\,x(-z(k))$$

$$\left. - z(k)\,\sqrt{\pi}\,\mathrm{erfc}\,x(z(k))\right)^{-1},$$

$$\left(2z(k)\,\sqrt{2}\left(2 - \frac{\sigma_p^2(k)}{\alpha^2}\right)\right.$$

$$+ \sqrt{\frac{\pi}{2}}\left(1 - \frac{\sigma_p^2(k)\,z^2(k)}{\alpha^2}\right)$$

$$\left.\times \left(\mathrm{erfc}\,x(z_p(k)) - \mathrm{erfc}\,x(-z_p(k))\right)\right)$$

$$\times \left(\sqrt{\frac{1}{\alpha^2} + \frac{1}{\sigma_p^2(k)}}\right.$$

$$\times \left(2 + z_p(k)\,\sqrt{\pi}\,\mathrm{erfc}\,x\left(-z_p(k)\right)\right.$$

$$\left.\left. - z_p(k)\,\sqrt{\pi}\,\mathrm{erfc}\,x\left(z_p(k)\right)\right)^{-1}\right). \tag{23}$$

Suppose that the input noise variance is known. To implement (20) or (23), we must know the parameter of the prior $\sigma(k)$ (suppose that $\sigma(k) = \sigma_p(k)$). Mihçak et al. [43] showed that using local variance (instead of global variance) for Wiener filter leads to a substantial improvement in denoising results (using local variance allows incorporating the local statistics of image into the proposed prior). It has been shown in the literature that the correctness of estimation of variance is an impact factor for denoising [23, 27, 34, 42–46]. Thus, the proposed criteria for estimation of the variance, such as the involved data for estimation (e.g., in some approaches the coarser scales are used as a source of prior), the type of estimator, and the shape and size of the proposed window for the local estimation of the variance, play key roles in the performance of denoising procedure. For example, in [54] a recurrence equation using a local Gaussian pdf is used for estimation of $\sigma(k)$ or in [64] the variable size of the locally adaptive window is obtained using a region-based approach. However, for each data point $\overline{y}(k)$, a simple estimation of

$\sigma(k)$ can be formed based on a local neighborhood $N(k)$. In simplest case, we can use a square window $N(k)$ centered at $\overline{y}(k)$ and suppose that in this window the variance is approximately constant. Then, an empirical estimate for $\sigma(k)$ can be obtained as follows:

$$\hat{\sigma}^2(k) = \frac{1}{2M} \sum_{j \in N(k)} \left(y^2(j) + y_p^2(j) \right) - \sigma_n^2, \qquad (24)$$

where M is the number of coefficients in $N(k)$ and σ_n can be estimated by [4] $\sigma_n = \text{median}\{|\text{noisy wavelet coefficients in finest scale}|\}/0.6745$. In this estimation, we propose the coarser scale as a source of prior, but another estimate can be obtained using only spatial adjacent in the same scale. It has been shown [47] that the local features in the edges of images are not isotropic and so can be better modeled in a shape-adaptive window selection manner. We explain in this regard in Section 4 and try to improve the denoising results by using anisotropic window instead of square window for the estimation of local parameters (such as variance in (24)).

3.2. Denoising Based on Modeling Noise-Free Data by a Mixture of Bivariate Gaussian PDFs with Local Parameters. A nonlinear shrinkage function for wavelet-based denoising, which is derived by assuming that the noise-free wavelet coefficients follow a bivariate Gaussian mixture model with local parameters given by (1), is introduced in this section. Substituting (1) in (17), we can write

$$\hat{w}(k) = \left(\iint w(k) \, p_{\overline{n}}(\overline{y}(k) - \overline{w}(k)) \right.$$
$$\times \left[a(k) \, p_1(\overline{w}(k)) \right.$$
$$\left. + (1 - a(k)) \, p_2(\overline{w}(k)) \right] d\overline{w}(k) \bigg)$$
$$\times \left(\iint p_{\overline{n}}(\overline{y}(k) - \overline{w}(k)) \right.$$
$$\times \left[a(k) \, p_1(\overline{w}(k)) \right.$$
$$\left. + (1 - a(k)) \, p_2(\overline{w}(k)) \right] d\overline{w}(k) \bigg)^{-1}$$
$$= \frac{a(k) \iint w(k) \, p_{\overline{n}}(\overline{y}(k) - \overline{w}(k)) \, p_1(\overline{w}(k)) \, d\overline{w}(k)}{a(k) \, g_1(\overline{y}(k)) + (1 - a(k)) \, g_2(\overline{y}(k))}$$
$$+ \bigg((1 - a(k))$$
$$\times \iint w(k) \, p_{\overline{n}}(\overline{y}(k) - \overline{w}(k)) \, p_2(\overline{w}(k)) \, d\overline{w}(k) \bigg)$$
$$\times \left(a(k) \, g_1(\overline{y}(k)) + (1 - a(k)) \, g_2(\overline{y}(k)) \right)^{-1}, \qquad (25)$$

where

$$g_i(\overline{y}(k)) = \iint p_{\overline{n}}(\overline{y}(k) - \overline{w}(k)) \, p_i(\overline{w}(k)) \, d\overline{w}(k), \qquad (26)$$
$$i = 1, 2.$$

In fact, $g_i(\overline{y}(k))$ is the 2D convolution of the pdf of $p_{\overline{n}}$ (defined in (15) or (16)) and p_i (defined in (1)). Using (15) as proposed model for noise, both $p_{\overline{n}}$ and p_i are bivariate Gaussian pdfs. So, we obtain

$$g_i(\overline{y}(k)) = \exp\bigg(-(1/2)$$
$$\times \left(y^2(k) / \left(\sigma_n^2 + \sigma_{i1}^2(k) \right) \right.$$
$$\left. + y_p^2(k) / \left(\sigma_n^2 + \sigma_{i2}^2(k) \right) \right) \bigg) \qquad (27)$$
$$\times \left(2\pi \sqrt{\left(\sigma_n^2 + \sigma_{i1}^2(k) \right) \left(\sigma_n^2 + \sigma_{i2}^2(k) \right)} \right)^{-1}$$
$$i = 1, 2.$$

For two-sided Rayleigh noise (16), more computations are needed. After some simplifications, the final formula would be

$$g_i(\overline{y}(k))$$
$$= \frac{\exp\left(-y^2(k) / 2\sigma_{i1}^2(k) - y_p^2(k) / 2\sigma_{i2}^2(k) \right)}{8\pi \left(1 + \sigma_{i1}^2(k) / \alpha^2 \right) \left(1 + \sigma_{i2}^2(k) / \alpha^2 \right) \sigma_{i1}(k) \, \sigma_{i2}(k)}$$
$$\times \left(2 + z_i(k) \, \sqrt{\pi} \, \text{erfc} \, x\left(-z_i(k) \right) \right.$$
$$\left. - z_i(k) \, \sqrt{\pi} \, \text{erfc} \, x\left(z_i(k) \right) \right)$$
$$\times \left(2 + z_{ip}(k) \, \sqrt{\pi} \, \text{erfc} \, x\left(-z_{ip}(k) \right) \right.$$
$$\left. - z_{ip}(k) \, \sqrt{\pi} \, \text{erfc} \, x\left(z_{ip}(k) \right) \right),$$
$$i = 1, 2, \qquad (28)$$

where

$$z_i(k) = \frac{y(k)}{\sigma_{i1}^2(k)} \sqrt{\frac{1}{2/\alpha^2 + (2/\sigma_{i1}^2(k))}}, \quad i = 1, 2,$$
$$\qquad (29)$$
$$z_{ip}(k) = \frac{y_p(k)}{\sigma_{i2}^2(k)} \sqrt{\frac{1}{2/\alpha^2 + (2/\sigma_{i2}^2(k))}}, \quad i = 1, 2.$$

Using (19), we can obtain numerators of (25), and finally (25) for AWGN can be written as

$$\hat{w}(k) = \left(\sigma_{11}^2(k) / \left(\sigma_{11}^2(k) + \sigma_n^2 \right) \right.$$
$$\left. + R(\overline{y}(k)) \left(\sigma_{21}^2(k) / \left(\sigma_{21}^2(k) + \sigma_n^2 \right) \right) \right) \qquad (30)$$
$$\times \left(1 + R(\overline{y}(k)) \right)^{-1} y(k),$$
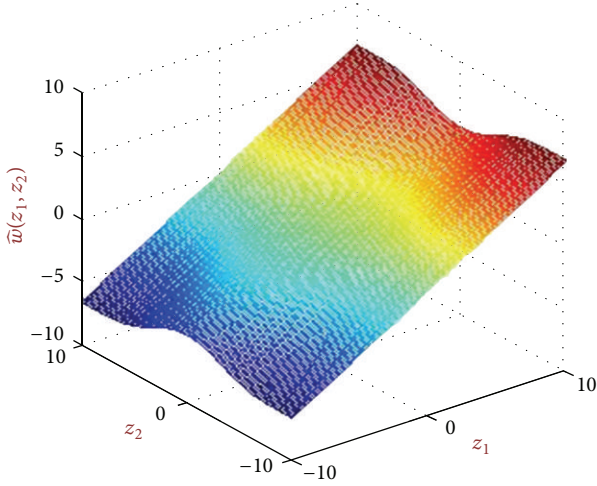
FIGURE 4: A shrinkage function produced from BiGaussMixShrink for sample parameters.

where

$$R\left(\overline{y}\left(k\right)\right) = \left(\left(\left(1 - a\left(k\right)\right)\right.\right.$$

$$\times \exp\left(-\frac{1}{2}\right.$$

$$\times \left(\frac{y^2\left(k\right)}{\sigma_n^2 + \sigma_{21}^2\left(k\right)} + \frac{y_p^2\left(k\right)}{\sigma_n^2 + \sigma_{22}^2\left(k\right)}\right)\right)\right)$$

$$\times \left(\sqrt{\left(\sigma_n^2 + \sigma_{21}^2\left(k\right)\right)\left(\sigma_n^2 + \sigma_{22}^2\left(k\right)\right)}\right)^{-1}\right)$$

$$\times \left(\left(\left(a\left(k\right)\exp\left(-\frac{1}{2}\left(\frac{y^2\left(k\right)}{\sigma_n^2 + \sigma_{11}^2\left(k\right)}\right.\right.\right.\right.$$

$$\left.\left.\left.\left.+\frac{y_p^2\left(k\right)}{\sigma_n^2 + \sigma_{12}^2\left(k\right)}\right)\right)\right)\right.$$

$$\times \left(\sqrt{\left(\sigma_n^2 + \sigma_{11}^2\left(k\right)\right)\left(\sigma_n^2 + \sigma_{12}^2\left(k\right)\right)}\right)^{-1}\right)^{-1}.$$

(31)

We call the new obtained bivariate local shrinkage function as *BiGaussMixShrinkL*. Figure 4 shows this shrinkage function with sample constant parameters.

Similarly, using (21), we can obtain numerators of (25), and finally (25) for two-sided Rayleigh noise is obtained as

$$\widehat{w}\left(k\right) = \frac{1}{1 + R\left(\overline{y}\left(k\right)\right)}$$

$$\times \left(2z_1\left(k\right)\sqrt{2}\left(2 - \frac{\sigma_{11}^2\left(k\right)}{\alpha^2}\right)\right.$$

$$+ \sqrt{\frac{\pi}{2}}\left(1 - \frac{\sigma_{11}^2\left(k\right)z_1^2\left(k\right)}{\alpha^2}\right)$$

$$\times \left(\text{erfc}\,x\left(z_1\left(k\right)\right) - \text{erfc}\,x\left(-z_1\left(k\right)\right)\right)\right)$$

$$\times \left(\sqrt{\frac{1}{\alpha^2} + \frac{1}{\sigma_{11}^2\left(k\right)}}\right.$$

$$\times \left(2 + z_1\left(k\right)\sqrt{\pi}\,\text{erfc}\,x\left(-z_1\left(k\right)\right)\right.$$

$$\left.-z_1\left(k\right)\sqrt{\pi}\,\text{erfc}\,x\left(z_1\left(k\right)\right)\right)\right)^{-1}$$

$$+ \frac{R\left(\overline{y}\left(k\right)\right)}{1 + R\left(\overline{y}\left(k\right)\right)}$$

$$\times \left(2z_2\left(k\right)\sqrt{2}\left(2 - \frac{\sigma_{21}^2\left(k\right)}{\alpha^2}\right)\right.$$

$$+ \sqrt{\frac{\pi}{2}}\left(1 - \frac{\sigma_{21}^2\left(k\right)z_2^2\left(k\right)}{\alpha^2}\right)$$

$$\times \left(\text{erfc}\,x\left(z_2\left(k\right)\right) - \text{erfc}\,x\left(-z_2\left(k\right)\right)\right)\right)$$

$$\times \left(\sqrt{\frac{1}{\alpha^2} + \frac{1}{\sigma_{21}^2\left(k\right)}}\right.$$

$$\times \left(2 + z_2\left(k\right)\sqrt{\pi}\,\text{erfc}\,x\left(-z_2\left(k\right)\right)\right.$$

$$\left.-z_2\left(k\right)\sqrt{\pi}\,\text{erfc}\,x\left(z_2\left(k\right)\right)\right)\right)^{-1},$$

(32)

where

$$R\left(\overline{y}\left(k\right)\right)$$

$$= \frac{1 - a\left(k\right)}{a\left(k\right)}$$

$$\times \frac{\left(1 + \sigma_{11}^2\left(k\right)/\alpha^2\right)\left(1 + \sigma_{12}^2\left(k\right)/\alpha^2\right)\sigma_{11}\left(k\right)\sigma_{12}\left(k\right)}{\left(1 + \sigma_{21}^2\left(k\right)/\alpha^2\right)\left(1 + \sigma_{22}^2\left(k\right)/\alpha^2\right)\sigma_{2i1}\left(k\right)\sigma_{22}\left(k\right)}$$

$$\times \frac{\exp\left(-y^2\left(k\right)/2\sigma_{21}^2\left(k\right) - y_p^2\left(k\right)/2\sigma_{22}^2\left(k\right)\right)}{\exp\left(-y^2\left(k\right)/2\sigma_{11}^2\left(k\right) - y_p^2\left(k\right)/2\sigma_{12}^2\left(k\right)\right)}$$

$$\times \left(\left(2 + z_2\left(k\right)\sqrt{\pi}\,\text{erfc}\,x\left(-z_2\left(k\right)\right)\right.\right.$$

$$\left.-z_2\left(k\right)\sqrt{\pi}\,\text{erfc}\,x\left(z_2\left(k\right)\right)\right)$$

$$\times \left(2 + z_{2p}\left(k\right)\sqrt{\pi}\,\text{erfc}\,x\left(-z_{2p}\left(k\right)\right)\right.$$

$$\left.\left.-z_{2p}\left(k\right)\sqrt{\pi}\,\text{erfc}\,x\left(z_{2p}\left(k\right)\right)\right)\right)$$

$$\times \left(\left(2 + z_1\left(k\right)\sqrt{\pi}\,\text{erfc}\,x\left(-z_1\left(k\right)\right)\right.\right.$$

$$\left.- z_1\left(k\right)\sqrt{\pi}\,\text{erfc}\,x\left(z_1\left(k\right)\right)\right)$$

$$\times \left(2 + z_{1p}(k)\,\sqrt{\pi}\,\mathrm{erfc}\,x\left(-z_{1p}(k)\right)\right.$$

$$\left.-z_{1p}(k)\,\sqrt{\pi}\,\mathrm{erfc}\,x\left(z_{1p}(k)\right)\right)\right)^{-1}.$$

$$(33)$$

We call this bivariate local shrinkage function as *BiGauss-RayMixShrinkL*. Figure 5 shows this shrinkage function with sample constant parameters.

For implementation of our denoising algorithm, we must estimate the parameters $\sigma_{ij}(k)$ for $i, j = 1, 2$, and $a(k)$ (that are for noise-free data) from noisy observation. For AWGN, the noisy observation would be a Gaussian mixture model with parameters $a(k), \sqrt{\sigma_n^2 + \sigma_{ij}^2(k)}$ for $i, j = 1, 2$. So, the following local EM algorithm is used to obtain the parameters.

E-step

$$r_1(k) \longleftarrow \frac{a(k)\,g_1\left(\overline{y}(k)\right)}{a(k)\,g_1\left(\overline{y}(k)\right) + (1 - a(k))\,g_2\left(\overline{y}(k)\right)},$$

$$r_2(k) \longleftarrow 1 - r_1(k).$$

$$(34)$$

M-step

$$a(k) \longleftarrow \frac{1}{M}\sum_{j \in N(K)} r_1(j),$$

$$(35)$$

$$\sigma_{1m}^2(k) \longleftarrow \frac{\sum_{j \in N(K)} r_i(j)\,y^2(k)}{\sum_{j \in N(K)} r_i(j)} - \sigma_n^2, \quad m = 1, 2,$$

$$(36)$$

$$\sigma_{2m}^2(k) \longleftarrow \frac{\sum_{j \in N(K)} r_i(j)\,y_p^2(k)}{\sum_{j \in N(K)} r_i(j)} - \sigma_n^2, \quad m = 1, 2,$$

$$(37)$$

where M is the number of coefficients in the window $N(K)$ centered at $\overline{y}(k)$. As discussed in the literatures [40], for non-Gaussian mixture models, which is a case for two-sided Rayleigh noise, using (34)–(36) finally converge to the final results.

Our denoising algorithm is summarized in Algorithm 1.

4. Shape Adaptive Windows Selection

It has been shown that using anisotropic and shape adaptive window for local parameter estimation can extremely improve the modeling and processing results. For example, in [47] a new image denoising is introduced that proposes an anisotropic window around each pixel of image and obtains the denoised pixel just by using the located data in the window. Comparing with the denoising methods that are based on proposing isotropic window around each pixel (e.g., [23, 27, 34, 42–46]), the proposed method in [47] is able to segment the image to rather smoothed regions before denoising due to anisotropic window selection that leads to improvement of denoising results. As explained before, the mixture model parameters in each subbands are estimated locally using an isotropic window around each
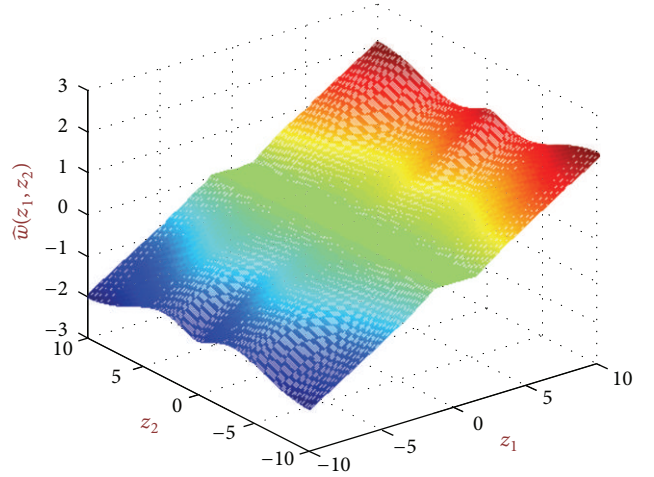

FIGURE 5: A shrinkage function produced from BiGaussRayMixShrink for sample parameters.

voxel. In this section at first we explain the structure of macular OCT then we introduce 3D "linear polynomial approximation-intersection confidence interval" (LPA-ICI) method for applying shape adaptive window selection around each voxel in 3D DCWT domain. So we will try the despeckling results in 3D DCWT domain by choosing an anisotropic window (instead of isotropic) for estimating the parameters of mixture model in each subband locally.

4.1. OCT Structure. To select the shape-adaptive window, we must take a look at the special structure of OCT data. In ophthalmology, the OCT data shows detailed images from within the retina. The automated analysis of OCT images can be used for the image-guided retinal therapy. Every year, many people become blind as a result of age-related macular degeneration (AMD) due to affecting the central retina where our central vision is perceived. The most sight-threatening form of AMD is called exudative or wet AMD. Choroidal neovascularization (CNV) is a common symptom of the degenerative maculopathy wet AMD. A wealth of powerful new treatments for CNV, especially anti-VEGF agents, have become available very recently to restore normal visual function. The risk of ocular adverse events, including the devastating intraocular infection, endophthalmitis, increases with repeated intravitreal treatment injections, and the effects of chronic treatment with anti-VEGF agents on the retina are unknown. Ideally a more cost-effective, patient-specific dosing strategy with the minimally necessary number of anti-VEGF injections is required. With all the promise, these novel treatments will only reach their full potential when objective and early indices of treatment response are developed. Prior to the introduction of retinal OCT imaging, clinical assessment of whether the preservation or restoration of visual function is successful, which indeed is the ultimate goal of treatment, could only be obtained by measuring visual function. Unfortunately, visual function lags structural response and is cumbersome and noisy, and its reproducibility is limited. Two-dimensional OCT imaging of the retina was introduced

Summarization of the proposed denoising algorithm

Step 1. Complex Wavelet Transformation of Noisy Image to Find $\overline{y}(k)$ and σ_n.

Step 2. Estimation of the Prior (Finding Mixture Model Parameters in Each Subband).

   2.1. Initialization for $a(k)$ and $\sigma_{11}(k), \sigma_{12}(k), \sigma_{21}(k), \sigma_{22}(k)$.

   2.2. Using (34) to find $r_1(k), r_2(k)$.

   2.3. Using (37) to update $a(k)$ by substituting $r_1(k), r_2(k)$ from Step 2.2.

   2.4. Updating the parameters of prior, $\sigma_{11}(k), \sigma_{12}(k), \sigma_{21}(k), \sigma_{22}(k)$, using (36) and (37).

   2.5. Finding $g_i(\overline{y}(k))$ from (27) for AWGN and (28) for two-sided Rayleigh noise using updated value in Step 2.3.

   2.6. Iteration Step 2.2 to 2.4 until parameter convergence.

Step 3. Substituting the Final Parameters in Step 2 in Shrinkage Function (30) for AWGN (after obtaining $R(\overline{y}(k))$ using (31)) and Shrinkage Function (32) for two-sided Rayleigh noise (after obtaining $R(\overline{y}(k))$ using (33)).

Step 4. Inverse Complex Wavelet Transformation.

ALGORITHM 1: Outline of the proposed denoising algorithm.
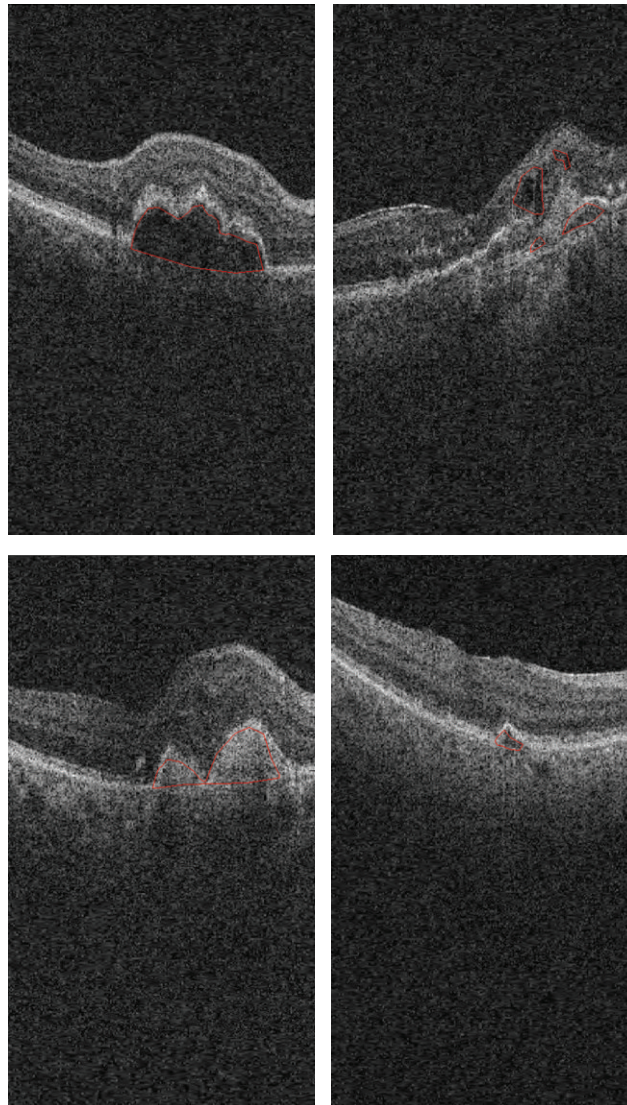
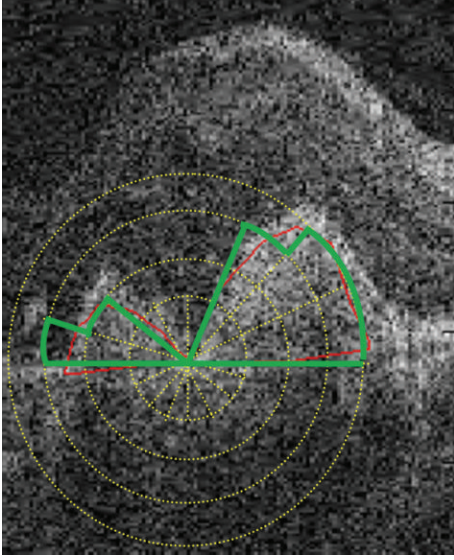FIGURE 6: Macular OCTs and detected SEADs by an expert.

FIGURE 7: The red line shows the detected SEAD by an expert. The yellow circles show the isotropic windows with various radii. The green line illustrates the obtained anisotropic based on LPA-ICI rule.

several years ago, and was rapidly adopted, among others, to qualitatively measure macular structure as an indicator of AMD treatment response and for guidance of retreatment in CNV recurrence. It is now becoming clear that these simplified structural measures though leading indicators of visual function are inadequate, as they are based on simplified interpretation of single transverse slices of the macula, some patients do not recover visual function even though their total macular thickness has become normally thin after treatment, and others paradoxically gain visual acuity while their macula is still thickened.

True 3D spectral OCT imaging, that became available in 2007 is fast (1.5 s per volume scan), allows full 3D retinal coverage at a much higher resolution and offers improved imaging of subtle differences in retinal structure. In the recent years [65, 66], 3D analysis of 3D OCT as an improved measure of subtle macular structure has been proposed motivated by various hypotheses as follows: *A model of retinal response to initial anti-VEGF treatment for CNV, based on quantitative 3D OCT-derived measures, can predict the timing of retreatment.*

On this base, developing analysis methods and approaches for 3D spectral OCT image analysis in the presence of wet AMD pathology (Symptomatic Exudate Associated Derangements or SEADs, also known as AMD-related cysts, vessel leakages, etc.) and assessing their performance by comparison to expert analyses are of utmost interest. Another interesting subject is determining how well the quantitative SEAD- and layer-derived measures from 3D OCT predict the patient-specific outcome parameters in response to postinduction anti-VEGF treatment in patients with CNV in order to predict the timing of retreatment.

Figure 6 shows several sample macular OCTs and detected SEADs by an expert as the region of interest (ROI). As
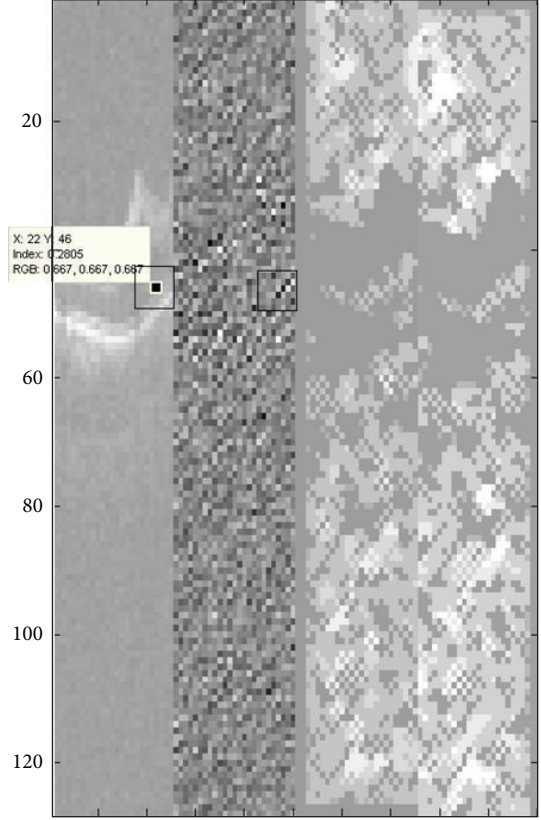

FIGURE 8: From left to right: imaginary LL subband of one slice of OCT data, the oriented (imaginary) subband around 45° (225°), $h^+(\cdot, 45°)$ for LL subband, and $h^+(\cdot, 225°)$ for LL subband extracted by applying LPA-ICI to the LL subband of imaginary part of DCWT. As indicated in the second image for $p_1 = (46, 20)$, $p_2 = (47, 21)$, and $p_3 = (46, 22)$ we would have $h^+(p_1, 45°) = 2$, $h^+(p_1, 225°) = 3$ (green dash), $h^+(p_2, 45°) = 1$, $h^+(p_2, 225°) = 3$ (orange dash), and $h^+(p_3, 45°) = 3$, $h^+(p_3, 225°) = 3$ (red dash).

we can see in this figure, the most important information of OCT data (about retina layers) is located in the center of OCT images.

4.2. 3D LPA-ICI for Data between the First and Last Layers. In [47], a new image denoising based on using an anisotropic window around each pixel of image is introduced. To select the anisotropic window, the linear directional filters $g_{h,\theta}$ that are obtained using local polynomial approximation (LPA) are employed. The θ indicates the direction of filter that is a member of countable set $\{\theta_1, \theta_2, \ldots, \theta_L\}$, where L is the number of directions. A common choice for L is $L = 8$ that results in the set $\{0°, 45°, 90°, 135°, 180°, 225°, 270°, 315°\}$. For each θ, the length of proposed window is selected from the countable and increasing set $\{h_1, h_2, \ldots, h_J\}$. So, for the noisy observation $y(k)$, we would have the following estimate:

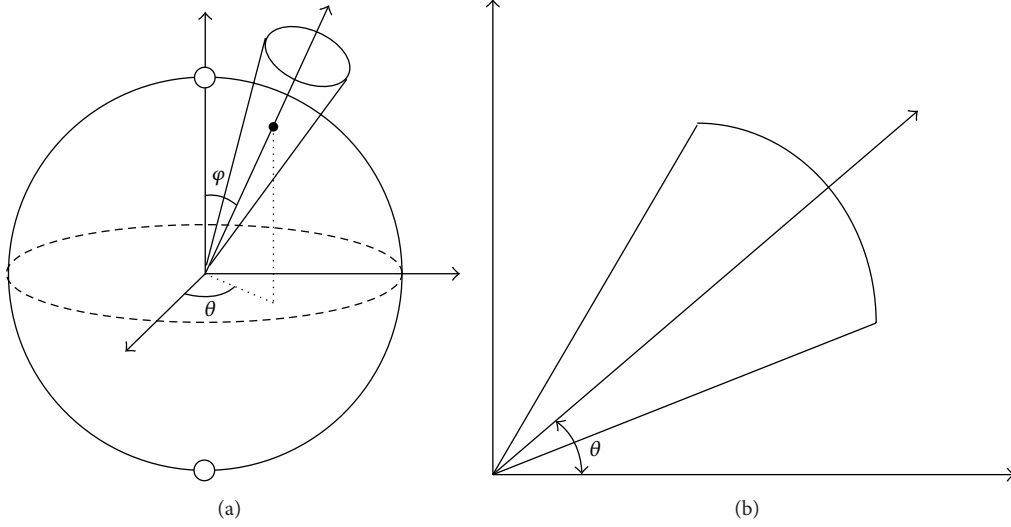$$x_{h,\theta}^{\text{est}}(k) = g_{h,\theta}(k) * y(k). \qquad (38)$$

FIGURE 9: Comparison between a circular sector for direction θ in 2D case (b) with a conical body produced for direction (θ, φ) in 3D case (a).
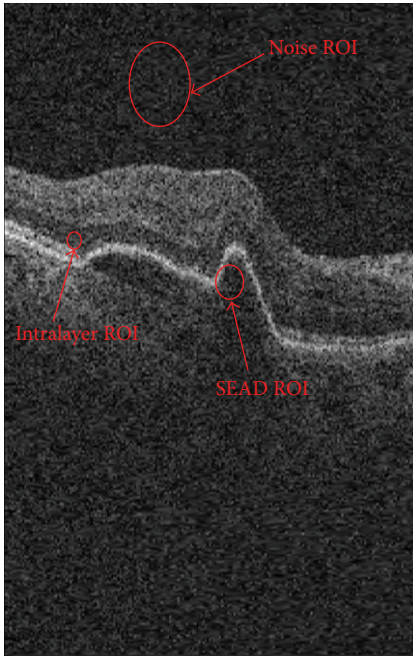

FIGURE 10: One slice from a sample OCT image and proposed ROIs for computation of MSNR and CNR reported in Table 1.

For each θ and k, an appropriate value of h called h^+ is estimated using the nonlinear intersection of confidence intervals (ICI) rule. h^+ is the largest h from the $h_1 < h_2 < \cdots < h_J$ provided that the estimated data using h^+ does not have noticeable difference with the estimated data with smaller h's. For this reason, the following confidence intervals are defined:

$$C_s = \left[x_{h_{s,\theta}}^{\text{est}}(k) - R\sigma_{x_{h_{s,\theta}}^{\text{est}}(k)}, \; x_{h_{s,\theta}}^{\text{est}}(k) + R\sigma_{x_{h_{s,\theta}}^{\text{est}}(k)} \right], \quad (39)$$

where R is the smoothing parameter and $\sigma_{x_{h_{s,\theta}}^{\text{est}}(k)}$ shows the variance of $x_{h_{s,\theta}}^{\text{est}}(k)$ and is obtained as follows:

$$\sigma_{x_{h_{s,\theta}}^{\text{est}}(k)}^2 = \int P_{x_{h_{s,\theta}}^{\text{est}}(k)} \int (f) \, df = \int P_y(f) G_{h,\theta}(f) \, df, \quad (40)$$

where $P(\cdot)$ shows the power spectral density function and $G_{h,\theta}(f)$ is the Fourier transform of $g_{h,\theta}$, and for a white random process, (40) is simplified to

$$\sigma_{x_{h_{s,\theta}}^{\text{est}}(k)}^2 = \int \sigma_n^2 G_{h,\theta}(f) \, df = \sigma_n^2 \sum g_{h,\theta}(k). \quad (41)$$

According to the ICI rule, D_s is defined using the following formula:

$$D_s = \bigcap_{i=1}^{s} C_i. \quad (42)$$

The largest s that leads to a nonempty value is called s^+, and finally $h^+(k,\theta)$ is obtained using $h^+(k,\theta) = h_{s^+}$.

Figure 7 shows an example of mentioned anisotropic window selection for a SEAD.

Since applying LPA-ICI in each subband is a time consuming process, a fast version of the mentioned algorithm can be based on only applying LPA-ICI to low-pass subbands using $L = 12$ with an offset of $15°$ that results in the set {$15°$, $45°$, $75°$, $105°$, $135°$, $165°$, $195°$, $225°$, $255°$, $285°$, $315°$, $345°$} in a 2D case. Since, in this case, each subband is extracting the information concentrated in a specific direction corresponding to {$15°$ ($195°$), $45°$ ($225°$), $75°$ ($255°$), $105°$ ($285°$), $135°$ ($315°$), $165°$ ($345°$)}, the extracted $h^+(k,\theta) = h_{s^+}$ that results from applying LPA-ICI to corresponding low-pass subband is used for obtaining the local parameters of kth pixel. For example, suppose that DCWT is used for 3 scales and we want to calculate the local parameters of coarsest scale for the oriented real subband around $45°$ ($225°$).

TABLE 1: The results of MSNR and CNR using several ROIs, shown in Figure 12.

Methods					
Local (L) Nonlocal (NL)	Homomorphic (H) Nonhomomorphic (NH)	Gaussian noise (G) Two-sided Rayleigh noise (R)	$\text{MSNR}_{\text{ROI1}}$	$\text{MSNR}_{\text{ROI2}}$	CNR
L	H	G	7.00	15.76	8.76
NL	H	G	7.56	17.03	9.47
L	NH	G	12.27	27.76	13.49
NL	NH	G	10.77	22.73	11.95
L	H	R	5.89	13.11	7.22
NL	H	R	8.63	19.59	10.95
L	NH	R	10.75	22.55	11.81
NL	NH	R	10.88	23.05	12.17
Original image			2.56	5.30	2.74

TABLE 2: PSNR (in dB) values of test images for different nonstationary noise levels.

Noise parameters $\sigma_g(i) = k_0 s(i) + k_1$	Lena				Boat				Barbara			
	Noisy Image	Soft thresh. [54]	Proposed method in [55]	Our method	Noisy image	Soft thresh. [54]	Proposed method in [55]	Our method	Noisy image	Soft thresh. [54]	Proposed method in [55]	Our method
$k_0 = 0.05, k_1 = 4$	27.72	34.13	34.61	**35.60**	27.51	32.44	32.59	**33.26**	27.94	31.99	32.20	**33.56**
$k_0 = 0.1, k_1 = 4$	23.48	31.62	32.49	**33.31**	23.22	29.95	30.23	**30.75**	23.69	29.05	25.81	**30.91**
$k_0 = 0.2, k_1 = 4$	18.49	27.50	29.65	**30.52**	18.20	26.77	27.60	**28.13**	18.71	25.81	25.94	**27.91**

TABLE 3: Comparison between PSNRs (in dB) of denoised images with Fast TI Haar algorithm [55] and BiGaussMixShrinkL.

	Lena	Boat	Barbara	Confocal Phantom	Shep Logan Phantom	Bowl
Noisy image	27.22	27.05	27.49	35.74	47.68	28.21
Fast TI Haar	32.11	29.30	26.59	44.49	60.63	46.79
BiGaussMixShrinL	**39.88**	**37.57**	**37.78**	**47.36**	**64.65**	**47.09**

For this reason, $h^+(k, 45°)$ and $h^+(k, 225°)$ are extracted from the results of applying LPA-ICI on the LL subband of real part (or imaginary part) of DCWT. Then, if we are in the jth scale, only $2^{j-1}h^+(k, 45°)$ pixel in direction of 45° and $2^{j-1}h^+(k, 225°)$ pixel in direction of 225° are used to extract the local parameters of kth pixel in this subband (Figure 8).

A similar manner can be proposed in 3D case [67]. However, instead of using 2D direction θ_i, we use 3D direction (θ_i, φ_i). As shown in Figure 9, in 2D case we use a circular sector for each direction while for 3D case a conical body is produced for direction (θ_i, φ_i), and the sphere is covered (partly) using these cones. Similar to 2D case $g_{h,\theta,\varphi}$, is defined and for each (θ_i, φ_i) the best h called $h^+(k, \theta_i, \varphi_i)$ is obtained using ICI rule.

Note that in order to incorporate the anisotropic window selection for each DCWT coefficient in our OCT denoising algorithm explained in Algorithm 1, instead of using a square window for parameter estimation, an anisotropic window is obtained for each coefficient k using the explained LPA-ICI method in this section, and only available data in this window are used for estimating $a(k)$ and $\sigma_{11}(k), \sigma_{12}(k), \sigma_{21}(k)$, and $\sigma_{22}(k)$.

5. Experimental Results

In this section, we apply the proposed despeckling algorithm to OCT image noise reduction. For this reason, we use 20 three-dimensional OCT datasets in the presence of wet AMD pathology (SEAD) and use mean signal-to-noise ratio (MSNR) and contrast-to-noise ratio (CNR) as two quality measurements for OCT data. To calculate these measurements, we must define the region of interest (ROI). In this paper, we propose this region within the SEAD as illustrated in Figure 10. The base MSNR and CNR are defined as follows:

$$\text{MSNR}_{\text{ROI}} = \frac{\mu_{\text{ROI}}}{\sigma},$$
$$\text{CNR} = |\text{MSNR}_{\text{ROI1}} - \text{MSNR}_{\text{ROI2}}|, \quad (43)$$

where μ_{ROI} shows the mean of ROI and σ indicates the standard deviation of a large region outside the ROI (noise ROI in Figure 10).

Table 1 shows the results of MSNR and CNR for proposed ROIs in OCT data using our algorithm. As discussed in Section 3, various shrinkage functions can be obtained
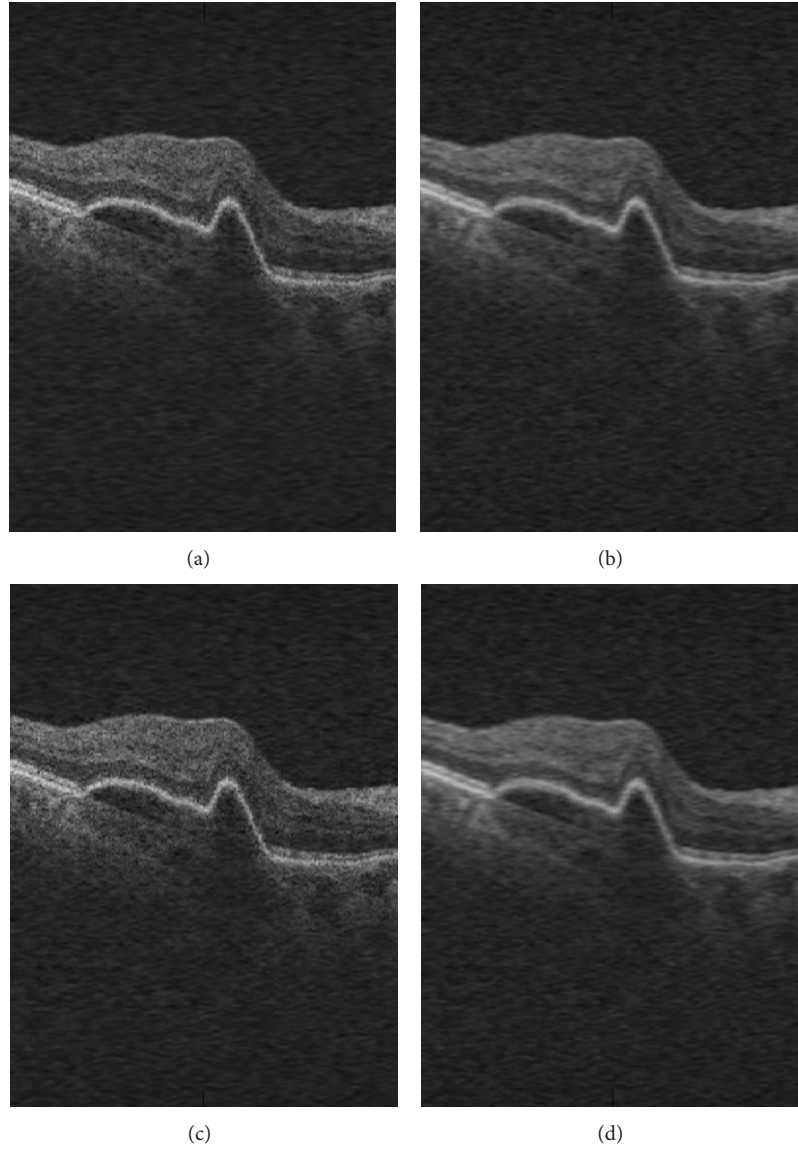
FIGURE 11: The results of applying homomorphic methods on proposed image in Figure 10. From top-left clockwise: despeckled data using *BiGaussMixShrinkL*, nonlocal *BiGaussMixShrinkL*, nonlocal *BiGaussRayMixShrinkL*, and *BiGaussRayMixShrinkL*.

using our algorithm based on applying log transformation before applying 3D DCWT (we use homomorphic prefix for this method and non-homomorphic when we do not use log transformation) and proposing AWGN or two-sided Rayleigh pdf for modeling noise in 3D DCWT domain (we name them *BiGaussMixShrinkL* and *BiGaussRayMixShrinkL*, resp.). Figures 11 and 12, respectively, show the results of applying non-homomorphic and homomorphic methods for (a slice of) depicted OCT image in Figure 10. In this figure, also in Table 1, we compare the results of nonlocal version of methods to show the effect of using anisotropic window selection technique. In order to show the SNR improvements, CNR curves for 156 selected ROIs have been depicted in Figure 13. It is clear that non-homomorphic *BiGaussMixShrinkL* method outperforms the others.

Another way for evaluating the effect of our despeckling algorithm is the investigation of the intralayer segmentation results. Figure 14 shows a comparison between the segmented layers of a 650 × 512 × 128 Topcon 3D OCT-1000 imaging system using proposed method in [53]. It is clear that the first layer is detected truly after despeckling.

6. Conclusion and Future Work

In this paper, we introduced a new noise reduction algorithm for 3D OCT data. We found new shrinkage functions employing a mixture of bivariate Gaussian for modeling wavelet coefficients in each subband of complex wavelets. The parameters of this mixture model are estimated locally using
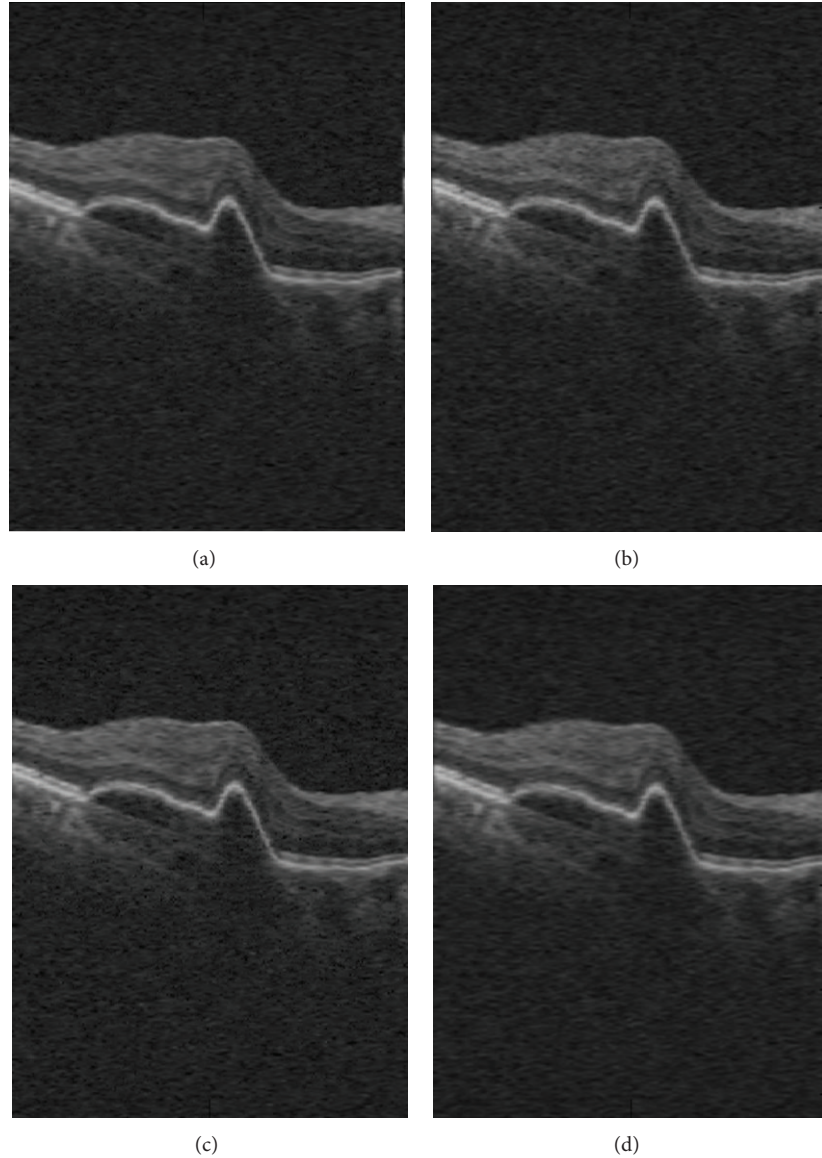
FIGURE 12: The results of applying non-homomorphic methods on proposed image in Figure 10. From top-left clockwise: despeckled data using *BiGaussMixShrinkL,* nonlocal *BiGaussMixShrinkL,* nonlocal *BiGaussRayMixShrinkL,* and *BiGaussRayMixShrinkL.*

a shape-adaptive manner based on the special structure of OCT data. We also used this model for denoising of other kinds of noise. Experiments show that our model has better results than other methods visually and in terms of PSNR especially for the crowded images. In this paper, we suppose that the parameters of EM algorithm, in extracted windows are constant. It is possible to improve the EM algorithm, for example, by using recurrence equations. It is possible that we only propose the main section of data (between the first and last layers) containing retina layer information and apply our algorithm on the selected data to improve the speed and performance of denoising process.

Using 3D DCWT instead of other transforms such as 3D DWT is a main reason for improvement of the denoising results [30]. In [27], it has been shown that other kinds of oriented transforms such as steerable pyramid decomposition can produce better denoising results. However, for 3D case, 3D transforms that are applied on whole 3D data (not slice by slice) such as surfacelet [68] and 3D discrete curvelet [69] can be investigated.

Appendices

A. Directional Selectivity Property of 2D DCWT

Since DWT in 2D domain is produced using separable (row-column) implementation, it has a poor directional selectivity.
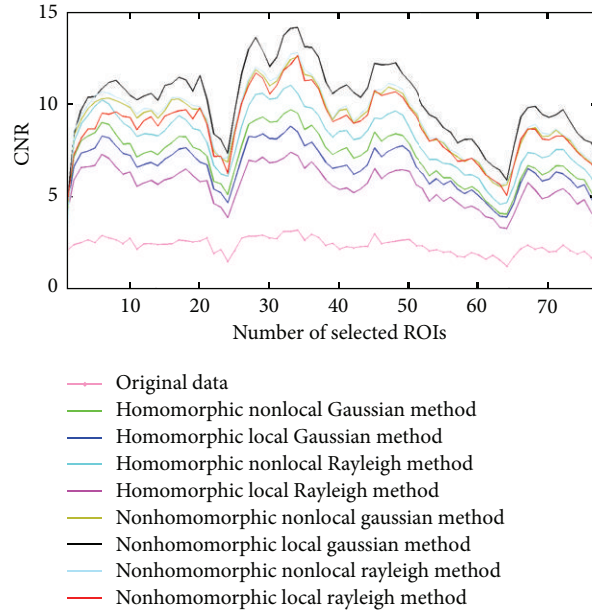
FIGURE 13: A comparison between CNR curves for 156 selected ROIs from OCT dataset.
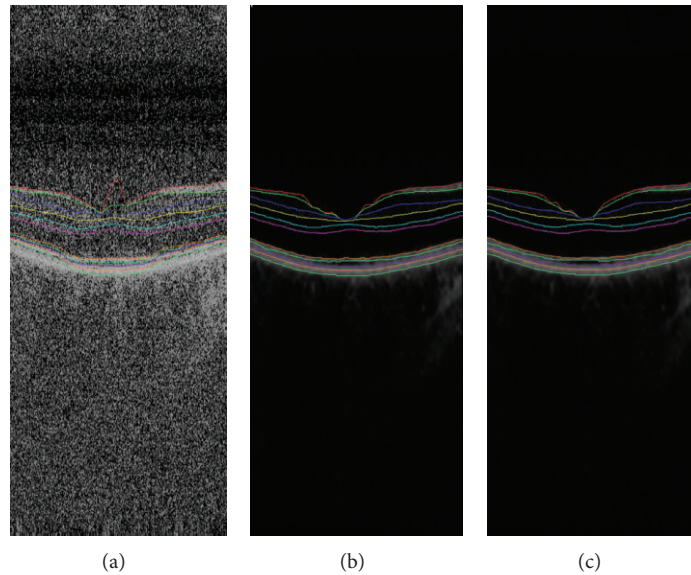

(a)　　　　　　　　(b)　　　　　　　　(c)

FIGURE 14: A comparison between the segmented layers of a $650 \times 512 \times 128$ Topcon 3D OCT-1000 imaging system using proposed method in [53]. From left to right: original image, denoised image by nonlocal homomorphic *BiGaussRayMixShrinkL* method*,* and local homomorphic *BiGaussRayMixShrinkL* method.

For example, the HH wavelet is the product of the high-pass functions along the first and second dimensions. Because DWT uses real filters, the HH wavelet mixes +45° and −45° orientations that results in the checkerboard artifact because it fails to isolate these orientations. In contrast, since the spectrum of the (approximately) analytic 1D wavelet is supported on only one side of the frequency axis, the spectrum of the DCWT in 2D domain is supported in only one quadrant of the 2D frequency plane. Figure 15 illustrates a comparison between subbands of 2D DWT and 2D DCWT.

B. A Sample Bivariate Gaussian Mixture Model

Figure 16 shows the pdf of a bivariate Gaussian mixture model for sample parameters. We can see the marginal distribution produced from the model in this figure.

C. Other Kinds of Noise

In this appendix, we briefly explain the abilities of the proposed denoising algorithm in this paper for other kinds of noise.
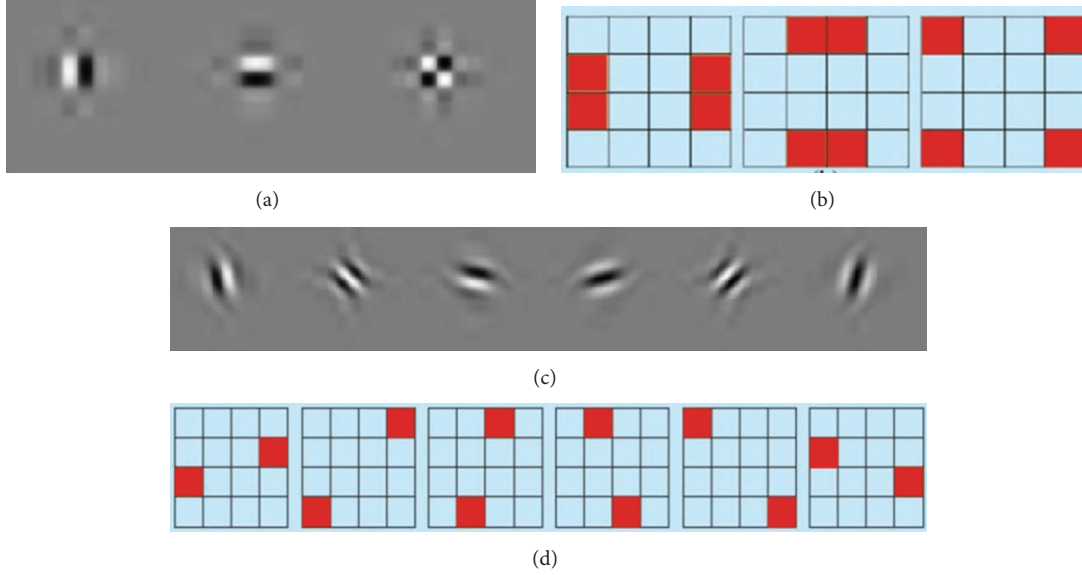
(a)

(b)

(c)

(d)

FIGURE 15: A comparison between subbands of DWT and DCWT. (a) The wavelets in the space domain (LH, HL, and HH). (b) The idealized support of the Fourier spectrum of each wavelet in the 2D frequency domain. We can see the checkerboard artifact of the third wavelet. (c) The complex wavelets in the space domain. (d) The idealized support of the Fourier spectrum of each wavelet in the 2D frequency plane. The absence of the checkerboard phenomenon is observed in both the space and frequency domains [48].
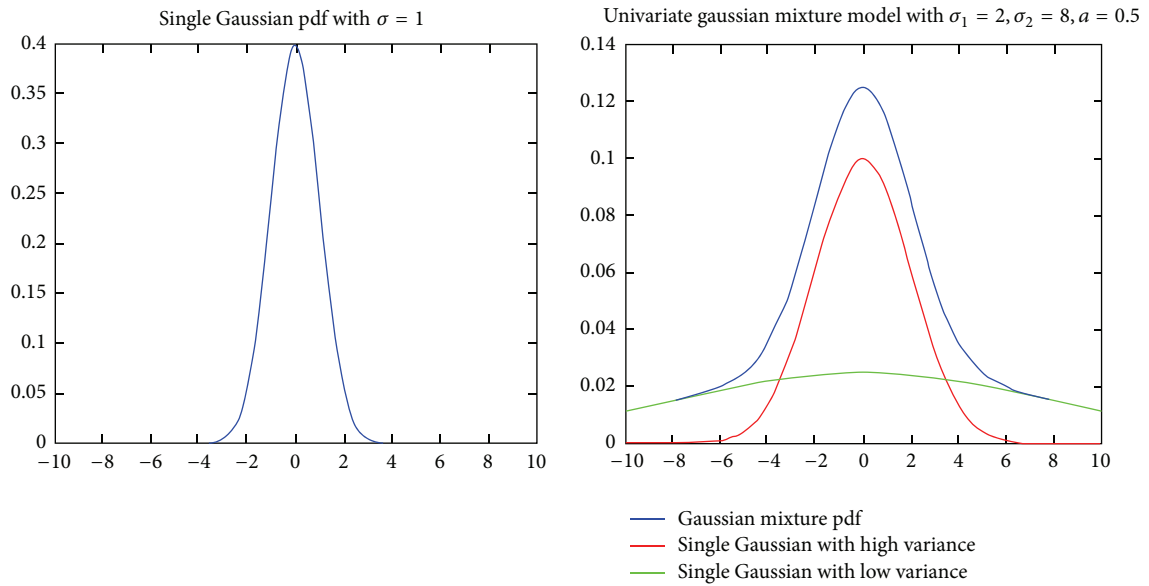
C.1. Stationary Noise. We tested the shrinkage function *BiGaussMixShrinkL* for stationary noise and compared it with other methods such as ProbShrink [34], BiShrink [45], and BLS-GSM [27] and found that our algorithm outperforms these techniques visually and in terms of peak signal-to-noise ratio (PSNR) for various levels of noise. For example, our algorithm for crowded images preserves details of images while BiShrink [45] in some cases produces blurred images (e.g., compare the area around the corresponding arrows in Figure 17). In addition, in [40] using other bivariate mixture models such as bivariate Laplacian was examined. We compared the results of using our model with method in [40], and our simulations show that our algorithm is faster and outperforms the reported shrinkage functions in [40] for some images. For example, for a 512×512 8-bit grayscale Barbara image, an improvement of 0.6 dB is obtained for noise level of 30, and *BiGaussMixShrinkL* is two times faster.

C.2. Nonstationary Noise. This section presents nonstationary noise reduction examples in complex wavelet domain. Although the stationary noise model is able to simplify the implementation of denoising algorithms, the statistical properties of the noise are not always accurately described with this assumption. For example, in some applications, the noise statistics are spatially varying and the noise power varies between pixels or samples. In these cases, the nonstationary noise assumption is more reasonable and can improve the denoising results. For example, we contaminate three 512×512 grayscale images, namely, Lena, Boat, and Barbara using signal-dependent Gaussian noise with variance $\sigma_g(i)$ that is defined as a linear function of the pixel intensities $s(i)$ as [54]:
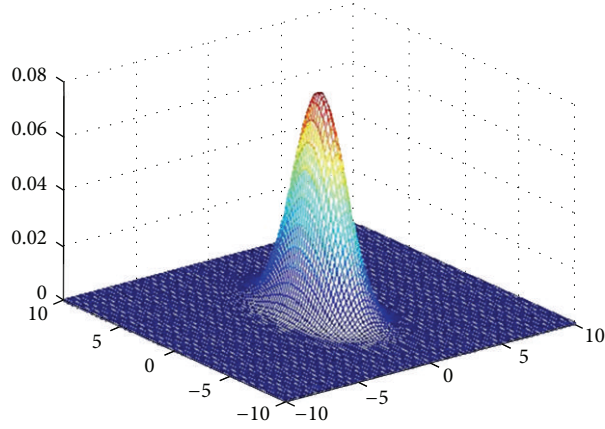
$$\sigma_g(i) = k_0 s(i) + k_1. \tag{C.1}$$

Since the variance of each noise component is spatially varying with the corresponding content of signal, the nonstationary processes are able to model the statistical properties of this noise. A comparison between the denoised image using soft thresholding, proposed method in [54], and the denoised image using a mixture of two bivariate Gaussian pdfs with local parameters (*BiGaussMixShrinkL*) for different noise levels can be seen in Table 2. In this table, the highest PSNR value is bolded. We can see from the table that our proposed algorithm has the better results compared to others especially for the Barbara image (which contains details) in the high-level noise.

In [54], it has been shown that the proposed algorithm based on the signal-dependent Gaussian noise can also be effective for the reduction of Poisson noise. On this base, we use our algorithm for noise reduction of images corrupted by Poisson noise generated using corresponding voxel intensities. For this reason, we use the software provided on http://willett.ece.wisc.edu/software.html to compare our method with Fast TI Haar algorithm [55]. The PSNR of grayscale images Lena, Boat, Barbara, Confocal Microscopy Phantom, Bowl, and Shepp Logan Phantom can be seen in Table 3. We can see that our algorithm outperforms the Fast TI Haar algorithm. Figure 18 illustrates a comparison between the denoised images produced from two algorithms. It is clear that our algorithm has better results especially for the crowded images. In fact, the Fast TI Haar algorithm has reasonable performance for soft images such as Confocal Microscopy Phantom, but since this algorithm blurs the produced images, for images with details such as Barbara, high-frequency features will be removed and we lose the important information. (LPA-ICI) method for applying shape adaptive window selection around each

Single Gaussian pdf with $\sigma = 1$

Univariate gaussian mixture model with $\sigma_1 = 2, \sigma_2 = 8, a = 0.5$

— Gaussian mixture pdf
— Single Gaussian with high variance
— Single Gaussian with low variance

Single bivariate Gaussian pdf with $\sigma_1 = 1, \sigma_2 = 2$

Bivariate Gaussian mixture model with $\sigma_{11} = 2, \sigma_{12} = 3, \sigma_{21} = 4, \sigma_{22} = 5, a = 5$
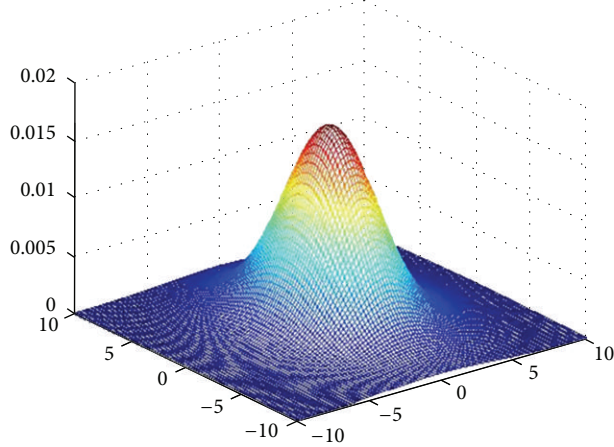
FIGURE 16: The pdf of a bivariate Gaussian mixture model for sample parameters and its marginal distribution.
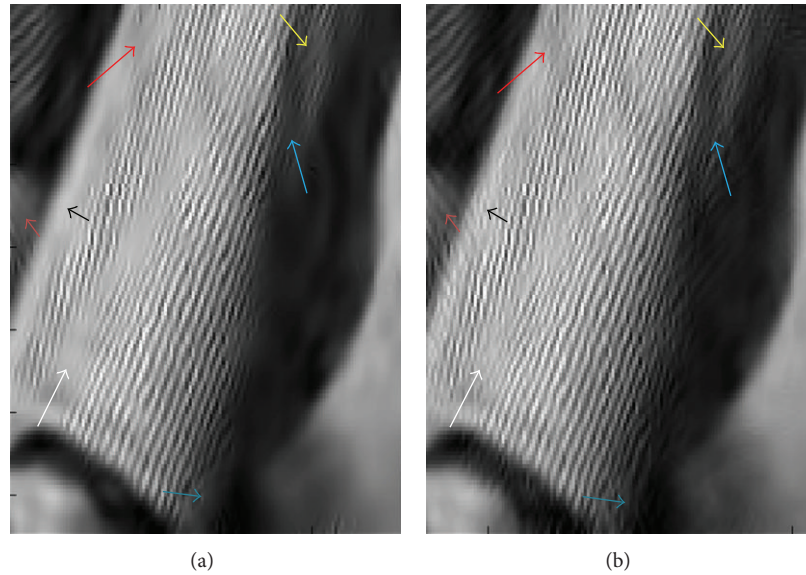
(a) (b)

FIGURE 17: (a) shows a part of Barbara image denoised using BiShrink [45] for stationary noise with $\sigma_n = 40$ and (b) shows denoised image using our method. Comparing the area around the corresponding arrows, we understand that our method is able to better preserve the details of images.
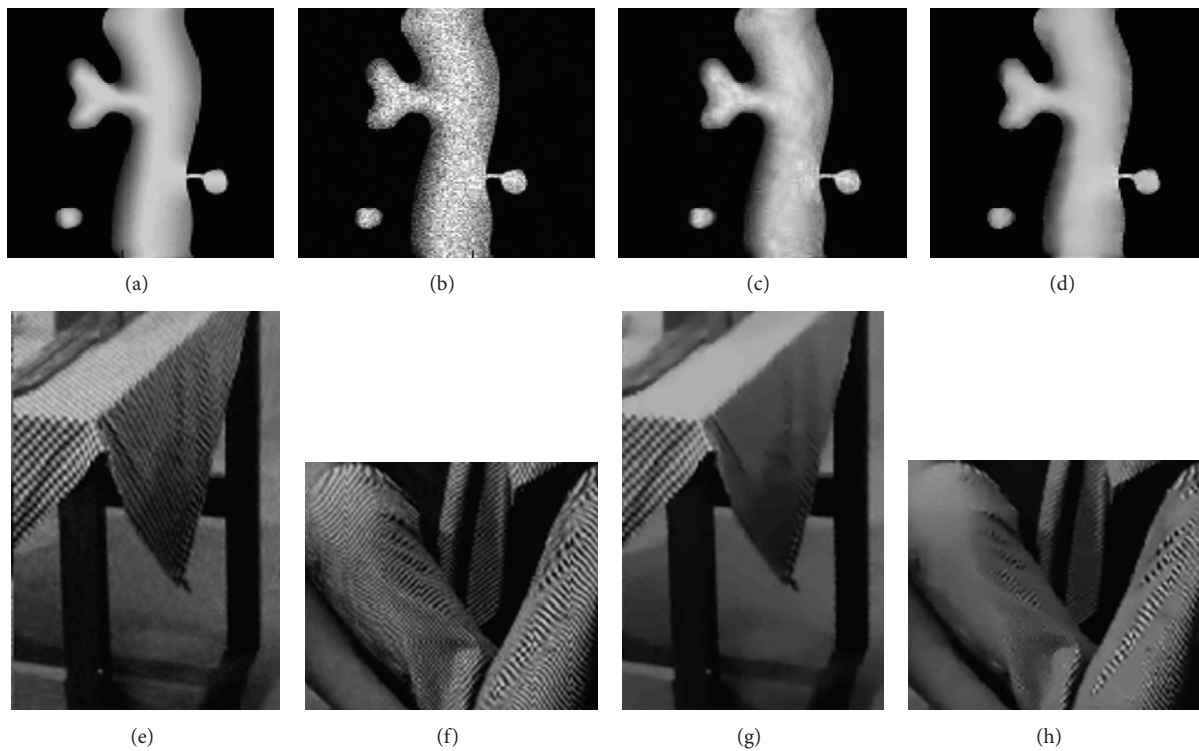

(a) (b) (c) (d)

(e) (f) (g) (h)

FIGURE 18: (a–d) show denoising results for Confocal Microscopy Phantom: from left to right: noise-free image, noisy image, and denoised image with our model and denoised image with Fast TI Haar algorithm. (e–h) show from left to right parts of denoised Barbara image with BiGaussMixShrinkL method and parts of denoised Barbara image with Fast TI Haar algorithm.

References

[1] D. Huang, E. A. Swanson, C. P. Lin et al., "Optical coherence tomography," *Science*, vol. 254, no. 5035, pp. 1178–1181, 1991.

[2] J. M. Schmitt, "Optical coherence tomography (OCT): a review," *IEEE Journal on Selected Topics in Quantum Electronics*, vol. 5, no. 4, pp. 1205–1215, 1999.

[3] M. A. Choma, M. V. Sarunic, C. Yang, and J. A. Izatt, "Sensitivity advantage of swept source and Fourier domain optical coherence tomography," *Optics Express*, vol. 11, no. 18, pp. 2183–2189, 2003.

[4] M. E. J. van Velthoven, D. J. Faber, F. D. Verbraak, T. G. van Leeuwen, and M. D. de Smet, "Recent developments in optical coherence tomography for imaging the retina," *Progress in Retinal and Eye Research*, vol. 26, no. 1, pp. 57–77, 2007.

[5] J. M. Schmitt, S. H. Xiang, and K. M. Yung, "Speckle in optical coherence tomography," *Journal of Biomedical Optics*, vol. 4, no. 1, pp. 95–105, 1999.

[6] D. L. Marks, T. S. Ralston, and S. A. Boppart, "Speckle reduction by I-divergence regularization in optical coherence tomography," *Journal of the Optical Society of America A*, vol. 22, no. 11, pp. 2366–2371, 2005.

[7] Z. Jian, Z. Yu, L. Yu, B. Rao, Z. Chen, and B. J. Tromberg, "Speckle attenuation in optical coherence tomography by curvelet shrinkage," *Optics Letters*, vol. 34, no. 10, pp. 1516–1518, 2009.

[8] Z. Jian, L. Yu, B. Rao, B. J. Tromberg, and Z. Chen, "Three-dimensional speckle suppression in optical coherence tomography based on the curvelet transform," *Optics Express*, vol. 18, no. 2, pp. 1024–1032, 2010.

[9] S. Chitchian, M. A. Fiddy, and N. M. Fried, "Denoising during optical coherence tomography of the prostate nerves via wavelet shrinkage using dual-tree complex wavelet transform," *Journal of Biomedical Optics*, vol. 14, no. 1, Article ID 014031, 2009.

[10] V. Zlokolica, Lj. Jovanov, A. Pizurica, and W. Philips, "Wavelet-based denoising for OCT images, Interaction between image processing, optics and photonics," in *Symposium on Optical Science and Technology*, Proceedings of SPIE, August 2007.

[11] V. Gupta, C. C. Chan, C.-L. Poh, T. H. Chow, T. C. Meng, and N. B. Koon, "Computerized automation of wavelet based denoising method to reduce speckle noise in OCT images," in *Proceedings of the 5th International Conference on Information Technology and Applications in Biomedicine*, pp. 120–123, May 2008.

[12] P. Puvanathasan and K. Bizheva, "Speckle noise reduction algorithm for optical coherence tomography based on interval type II fuzzy set," *Optics Express*, vol. 15, no. 24, pp. 15747–15758, 2007.

[13] A. Pižurica, L. Jovanov, B. Huysmans et al., "Multiresolution denoising for optical coherence tomography: a review and evaluation," *Current Medical Imaging Reviews*, vol. 4, no. 4, pp. 270–284, 2008.

[14] H. M. Salinas and D. C. Fernández, "Comparison of PDE-based nonlinear diffusion approaches for image enhancement and denoising in optical coherence tomography," *IEEE Transactions on Medical Imaging*, vol. 26, no. 6, pp. 761–771, 2007.

[15] K. M. Yung, S. L. Lee, and J. M. Schmilt, "Phase-domain processing of optical coherence tomography images," *Journal of Biomedical Optics*, vol. 4, no. 1, pp. 125–136, 1999.

[16] J. Rogowska and M. E. Brezinski, "Evaluation of the adaptive speckle suppression filter for coronary optical coherence tomography imaging," *IEEE Transactions on Medical Imaging*, vol. 19, no. 12, pp. 1261–1266, 2000.

[17] A. Ozcan, A. Bilenca, A. E. Desjardins, B. E. Bouma, and G. J. Tearney, "Speckle reduction in optical coherence tomography images using digital filtering," *Journal of the Optical Society of America A*, vol. 24, no. 7, pp. 1901–1910, 2007.

[18] D. C. Adler, T. H. Ko, and J. G. Fujimoto, "Speckle reduction in optical coherence tomography images by use of a spatially adaptive wavelet filter," *Optics Letters*, vol. 29, no. 24, pp. 2878–2880, 2004.

[19] S. G. Mallat, *A Wavelet Tour of Signal Processing*, Academic Press, San Diego, Calif, USA, 1998.

[20] B. Vidakovic, *Statistical Modeling by Wavelets*, John Wiley & Sons, New York, NY, USA, 1999.

[21] E. P. Simoncelli and E. H. Adelson, "Noise removal via Bayesian wavelet coring," in *Proceedings of the IEEE International Conference on Image Processing*, pp. 379–382, Lausanne, Switzerland, September 1996.

[22] E. P. Simoncelli, "Bayesian denoising of visual images in the wavelet domain," in *Bayesian Inference in Wavelet Based Models*, P. Muller and B. Vidakovic, Eds., pp. 291–308, Springer, New York, NY, USA, 1999.

[23] A. Achim, A. Bezerianos, and P. Tsakalides, "Novel Bayesian multiscale method for speckle removal in medical ultrasound images," *IEEE Transactions on Medical Imaging*, vol. 20, no. 8, pp. 772–783, 2001.

[24] J. M. Fadili and L. Boubchir, "Analytical form for a Bayesian wavelet estimator of images using the Bessel K form densities," *IEEE Transactions on Image Processing*, vol. 14, no. 2, pp. 231–240, 2005.

[25] P. A. Khazron and I. W. Selesnick, "Bayesian estimation of bessel K form random vectors in AWGN," *IEEE Signal Processing Letters*, vol. 15, pp. 261–264, 2008.

[26] M. S. Crouse, R. D. Nowak, and R. G. Baraniuk, "Wavelet-based statistical signal processing using hidden Markov models," *IEEE Transactions on Signal Processing*, vol. 46, no. 4, pp. 886–902, 1998.

[27] J. Portilla, V. Strela, M. J. Wainwright, and E. P. Simoncelli, "Image denoising using scale mixtures of Gaussians in the wavelet domain," *IEEE Transactions on Image Processing*, vol. 12, no. 11, pp. 1338–1351, 2003.

[28] H. Rabbani and S. Gazor, "Image denoising employing local mixture models in sparse domains," *IET Image Processing*, vol. 4, no. 5, pp. 413–428, 2010.

[29] A. Elmzoughi, A. Benazza-Benyahia, and J.-C. Pesquet, "An interscale multivariate statistical model for map multicomponent image denoising in the wavelet transform domain," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '05)*, pp. II45–II48, March 2005.

[30] H. Rabbani and M. Vafadust, "Image/video denoising based on a mixture of Laplace distributions with local parameters in multidimensional complex wavelet domain," *Signal Processing*, vol. 88, no. 1, pp. 158–173, 2008.

[31] H. A. Chipman, E. D. Kolaczyk, and R. E. McCulloch, "Adaptive Bayesian wavelet shrinkage," *Journal of the American Statistical Association*, vol. 92, no. 440, pp. 1413–1421, 1997.

[32] M. J. Borran and R. D. Nowak, "Wavelet-based denoising using hidden Markov models," in *Proceedings of the IEEE Interntional Conference on Acoustics, Speech, and Signal Processing*, pp. 3925–3928, May 2001.

[33] J. K. Romberg, H. Choi, and R. G. Baraniuk, "Bayesian tree-structured image modeling using wavelet-domain hidden Markov models," *IEEE Transactions on Image Processing*, vol. 10, no. 7, pp. 1056–1068, 2001.

[34] A. Pižurica, W. Philips, I. Lemahieu, and M. Acheroy, "A joint inter- and intrascale statistical model for Bayesian wavelet based image denoising," *IEEE Transactions on Image Processing*, vol. 11, no. 5, pp. 545–557, 2002.

[35] A. A. Bharath and J. Ng, "A steerable complex wavelet construction and its application to image denoising," *IEEE Transactions on Image Processing*, vol. 14, no. 7, pp. 948–959, 2005.

[36] L. Şendur and 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.

[37] H. Rabbani, M. Vafadust, S. Gazor, and I. Selesnick, "Image denoising employing a bivariate cauchy distribution with local variance in complex wavelet domain," in *Proceedings of the 12th IEEE Digital Signal Processing Workshop and 4th IEEE Signal Processing Education Workshop*, pp. 203–208, Grand Teton National Park, Wy, USA, September 2006.

[38] J. Portilla, "Full blind denoising through noise covariance estimation using gaussian scale mixtures in the wavelet domain," in *Proceedings of the International Conference on Image Processing (ICIP '04)*, pp. 1217–1220, October 2004.

[39] S. Lyu and E. P. Simoncelli, "Modeling multiscale subbands of photographic images with fields of Gaussian scale mixtures," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 4, pp. 693–706, 2009.

[40] H. Rabbani, R. Nezafat, and S. Gazor, "Wavelet-domain medical image denoising using bivariate laplacian mixture model," *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 12, pp. 2826–2837, 2009.

[41] I. Selesnick and K. Li, "Video denoising using 2d and 3d dual-tree complex wavelet transforms," in *Wavelet Applications in Signal and Image Processing*, vol. 5207 of *Proceedings of SPIE*, August 2003.

[42] 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.

[43] M. K. Mihçak, 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.

[44] H. Rabbani, M. Vafadust, P. Abolmaesumi, and S. Gazor, "Speckle noise reduction of medical ultrasound images in complex wavelet domain using mixture priors," *IEEE Transactions on Biomedical Engineering*, vol. 55, no. 9, pp. 2152–2160, 2008.

[45] L. Şendur and I. W. Selesnick, "Bivariate shrinkage with local variance estimation," *IEEE Signal Processing Letters*, vol. 9, no. 12, pp. 438–441, 2002.

[46] H. Rabbani, M. Vafadust, and S. Gazor, "Image denoising based on a mixture of bivariate gaussian distributions with local parameters in complex wavelet domain," in *Proceedings of the International Conference on Biomedical and Pharmaceutical Engineering*, pp. 174–179, December 2006.

[47] V. Katkovnik, K. Egiazarian, and J. Astola, "Adaptive window size image de-noising based on intersection of confidence intervals (ICI) rule," *Journal of Mathematical Imaging and Vision*, vol. 16, no. 3, pp. 223–235, 2002.

[48] I. W. Selesnick, R. G. Baraniuk, and N. G. Kingsbury, "The dual-tree complex wavelet transform," *IEEE Signal Processing Magazine*, vol. 22, no. 6, pp. 123–151, 2005.

[49] B. Wang, Y. Wang, I. Selesnick, and A. Vetro, "Video coding using 3D dual-tree wavelet transform," *Eurasip Journal on Image and Video Processing*, vol. 2007, Article ID 42761, 15 pages, 2007.

[50] W. Jingwei, G. Xinbo, and Z. Juanjuan, "A video watermarking based on 3-D complex wavelet," in *Proceedings of the 14th IEEE International Conference on Image Processing*, pp. V493–V496, September 2007.

[51] J. Yang, Y. Wang, W. Xu, and Q. Dai, "Image and video denoising using adaptive dual-tree discrete wavelet packets," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 19, no. 5, pp. 642–655, 2009.

[52] I. Bayram and I. W. Selesnick, "A dual-tree rational-dilation complex wavelet transform," *IEEE Transactions on Signal Processing*, vol. 59, no. 12, pp. 6251–6256, 2011.

[53] K. Lee, M. Niemeijer, M. K. Garvin, Y. H. Kwon, M. Sonka, and M. D. Abramoff, "Segmentation of the optic disc in 3-D OCT scans of the optic nerve head," *IEEE Transactions on Medical Imaging*, vol. 29, no. 1, pp. 159–168, 2010.

[54] W. Y. Lo and I. W. Selesnick, "Wavelet-domain soft-thresholding for non-stationary noise," in *Proceedings of the IEEE International Conference on Image Processing*, pp. 1441–1444, October 2006.

[55] R. M. Willett and R. D. Nowak, "Fast multiresolution photon-limited image reconstruction," in *Proceedings of the 2nd IEEE International Symposium on Biomedical Imaging: Macro to Nano (ISBI '04)*, vol. 2, pp. 1192–1195, April 2004.

[56] A. Pižurica, W. Philips, I. Lemahieu, and M. Acheroy, "A versatile wavelet domain noise filtration technique for medical imaging," *IEEE Transactions on Medical Imaging*, vol. 22, no. 3, pp. 323–331, 2003.

[57] S. Gupta, R. C. Chauhan, and S. C. Saxena, "Robust non-homomorphic approach for speckle reduction in medical ultrasound images," *Medical and Biological Engineering and Computing*, vol. 43, no. 2, pp. 189–195, 2005.

[58] S. Gupta, L. Kaur, R. C. Chauhan, and S. C. Saxena, "A versatile technique for visual enhancement of medical ultrasound images," *Digital Signal Processing*, vol. 17, no. 3, pp. 542–560, 2007.

[59] S. Yan, J. Yuan, M. Liu, and C. Hou, "Speckle noise reduction of ultrasound images based on an undecimated wavelet packet transform domain nonhomomorphic filtering," in *Proceedings of the 2nd International Conference on Biomedical Engineering and Informatics*, pp. 1–5, October 2009.

[60] R. F. Wagner, M. F. Insana, and D. G. Brown, "Statistical properties of radio-frequency and envelope-detected signals with applications to medical ultrasound," *Journal of the Optical Society of America A*, vol. 4, no. 5, pp. 910–922, 1987.

[61] P. M. Shankar, "Ultrasonic tissue characterization using a generalized Nakagami model," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 48, no. 6, pp. 1716–1720, 2001.

[62] P. M. Shankar, J. M. Reid, H. Ortega, C. W. Piccoli, and B. B. Goldberg, "Use of non-Rayleigh statistics for the identification of tumors in ultrasonic B-scans of the breast," *IEEE Transactions on Medical Imaging*, vol. 12, no. 4, pp. 687–692, 1993.

[63] P. M. Shankar, "A model for ultrasonic scattering from tissues based on the K distribution," *Physics in Medicine and Biology*, vol. 40, no. 10, pp. 1633–1649, 1995.

[64] I. K. Eom and Y. S. Kim, "Wavelet-based denoising with nearly arbitrarily shaped windows," *IEEE Signal Processing Letters*, vol. 11, no. 12, pp. 937–940, 2004.

[65] W. Drexler and J. G. Fujimoto, "State-of-the-art retinal optical coherence tomography," *Progress in Retinal and Eye Research*, vol. 27, no. 1, pp. 45–88, 2008.

[66] M. D. Abramoff, M. K. Garvin, and M. Sonka, "Retinal imaging and image analysis," *IEEE Reviews in Biomedical Engineering*, vol. 3, pp. 169–208, 2010.

[67] C. Ercole, A. Foi, V. Katkovnik, and K. Egiazarian, "Spatio-temporal pointwise adaptive denoising of video: 3D non-parametric approach," in *Proceedings of the 1st International Workshop on Video Process and Quality Metrics for Consumer Electronics (VPQM '05)*, 2005.

[68] Y. M. Lu and M. N. Do, "Multidimensional directional filter banks and surfacelets," *IEEE Transactions on Image Processing*, vol. 16, no. 4, pp. 918–931, 2007.

[69] L. Ying, L. Demanet, and E. Candès, "3D discrete curvelet transform," in *Wavelets XI*, vol. 5914 of *Proceedings of SPIE*, pp. 1–11, August 2005.