



# Statistical model for OCT image denoising

MUXINGZI LI, RAMZI IDOUGHI, BISWARUP CHOUDHURY, AND WOLFGANG HEIDRICH\*

King Abdullah University of science and Technology, Thuwal 23955-6900, Saudi Arabia

\*wolfgang.heidrich@kaust.edu.sa

<http://vccimaging.org>

**Abstract:** Optical coherence tomography (OCT) is a non-invasive technique with a large array of applications in clinical imaging and biological tissue visualization. However, the presence of speckle noise affects the analysis of OCT images and their diagnostic utility. In this article, we introduce a new OCT denoising algorithm. The proposed method is founded on a numerical optimization framework based on maximum-a-posteriori estimate of the noise-free OCT image. It combines a novel speckle noise model, derived from local statistics of empirical spectral domain OCT (SD-OCT) data, with a Huber variant of total variation regularization for edge preservation. The proposed approach exhibits satisfying results in terms of speckle noise reduction as well as edge preservation, at reduced computational cost.

© 2017 Optical Society of America

**OCIS codes:** (110.4500) Optical coherence tomography; (030.4280) Noise in imaging systems; (030.6140) Speckle; (030.6600) Statistical optics; (100.2980) Image enhancement.

## References and links

1. B. Karamata, K. Hassler, M. Laubscher, and T. Lasser, "Speckle statistics in optical coherence tomography," *J. Opt. Soc. Am. A* **22**(4), 593–596 (2005).
2. M. Kirillin, G. Farhat, E. Sergeeva, M. Kolios, and A. Vitkin, "Speckle statistics in OCT images: Monte Carlo simulations and experimental studies," *Opt. Lett.* **39**(12), 3472–3475 (2014).
3. D. Jesus and D. Iskander, "Assessment of corneal properties based on statistical modeling of OCT speckle," *Biomed. Opt. Express* **8**(1), 162–176 (2017).
4. M. Ali and B. Hadj, "Segmentation of OCT skin images by classification of speckle statistical parameters," in Proceedings of IEEE International Conference on Image Processing (IEEE, 2010), pp. 613–616.
5. A. Desjardins, B. Vakoc, W. Oh, S. Motaghianezam, G. Tearney, and B. Bouma, "Angle-resolved Optical Coherence Tomography with sequential angular selectivity for speckle reduction," *Opt. Express* **15**(10), 6200–6209 (2007).
6. D. Popescu, M. Hewko, M. Sowa, "Speckle noise attenuation in optical coherence tomography by compounding images acquired at different positions of the sample," *Opt. Commun.* **269**(1), 247–251 (2007).
7. A. Ozcan, A. Bilenca, A. E. Desjardins, B. E. Bouma, and G. J. Tearney, "Speckle reduction in optical coherence tomography images using digital filtering," *J. Opt. Soc. Am. A* **24**(7), 1901–1910 (2007).
8. M.A. Mayer, A. Borsdorf, M. Wagner, J. Hornegger, C.Y. Mardin and R.P. Tornow, "Wavelet denoising of multiframe optical coherence tomography data," *Biomed. Opt. Express* **3**(3), 572–589 (2012).
9. Z. Jian, L. Yu, B. Rao, B.J. Tromberg, and Z. Chen, "Three-dimensional speckle suppression in optical coherence tomography based on the curvete transform," *Opt. Express* **18**(2), 1024–1032 (2010).
10. L. Bian, J. Suo, F. Chen, and Q. Dai, "Multiframe denoising of high-speed optical coherence tomography data using interframe and intraframe priors," *J. Biomed. Opt.* **20**(3), 036006 (2015).
11. L. Fang, S. Li, Q. Nie, J. Izatt, C. A. Toth and S. Farsiu, "Sparsity based denoising of spectral domain optical coherence tomography images," *Biomed. Opt. Express* **3**(5), 927–942 (2012).
12. L. Fang, S. Li, R. P. McNabb, Q. Nie, A. N. Kuo, C. A. Toth, J. A. Izatt, and S. Farsiu, "Fast Acquisition and Reconstruction of Optical Coherence Tomography Images via Sparse Representation," *IEEE Trans. Med. Imag.* **32**(11), 2034–2049 (2013).
13. L. Fang, S. Li, D. Cunefare, and S. Farsiu, "Segmentation Based Sparse Reconstruction of Optical Coherence Tomography Images," *IEEE Trans. Med. Imag.* **36**(2), 407–421 (2017).
14. G. Gong, H. Zhang and M. Yao, "Speckle noise reduction algorithm with total variation regularization in optical coherence tomography," *Opt. Express* **23**(19), 24699–24712 (2015).
15. J. Duan, C. Tench, I. Gottlob, F. Proudlock, and L. Bai, "New variational image decomposition model for simultaneously denoising and segmenting optical coherence tomography images," *Phys. Med. Biol.* **60**(22), 8901–8922 (2015).
16. J. Duan, W. Lu, C. Tench, I. Gottlob, F. Proudlock, N. Samani, and L. Bai, "Denoising optical coherence tomography using second order total generalized variation decomposition," *Biomed. Signal Process. Control* **24**, 120–127 (2016).

17. H. M. Salinas, and D. C. Fernandez, "Comparison of PDE-based nonlinear diffusion approaches for image enhancement and denoising in Optical Coherence Tomography," *IEEE Trans. Med. Imag.* **26**(6), 761–771 (2007).
18. A. Wong, A. Mishra, K. Bizheva, and D.A. Clausi, "General Bayesian estimation for speckle noise reduction in optical coherence tomography retinal imagery," *Opt. Express* **18**(8), 8338–8352 (2010).
19. A. Cameron, D. Lui, A. Boroomand, J. Glaister, A. Wong, and K. Bizheva, "Stochastic speckle noise compensation in optical coherence tomography using non-stationary spline-based speckle noise modelling," *Biomed. Opt. Express* **4**(9), 1769–1785 (2013).
20. J. Aum, J. Kim, and J. Jeong, "Effective speckle noise suppression in optical coherence tomography images using nonlocal means denoising filter with double Gaussian anisotropic kernels," *Appl. Opt.* **54**(13), D43–D50 (2015).
21. J. Portilla, "Blind non-white noise removal in images using gaussian scale mixtures in the wavelet domain," in Proceedings of IEEE Benelux Signal Processing Symposium (IEEE, 2004), pp. 17–20.
22. P. Puvanathasan, and K. Bizheva, "Speckle noise reduction algorithm for optical coherence tomography based on interval type II fuzzy set," *Opt. Express* **15**(24), 15747–15758 (2007).
23. B. Chong, Y. Zhu, "Speckle reduction in optical coherence tomography images of human finger skin by wavelet modified {BM3D} filter," *Opt. Commun.* **291**, 461–469 (2013).
24. R. Kafieh, H. Rabbani, and I. Selesnick, "Three Dimensional Data-Driven Multi Scale Atomic Representation of Optical Coherence Tomography," *IEEE Trans. Med. Imag.* **34**(5), 1042–1062 (2015).
25. T. Mahmudi, R. Kafieh, and H. Rabbani, "Comparison of macular OCTs in right and left eyes of normal people," *Proc. SPIE* **9038**, 90381W (2014).
26. A. Chambolle, and T. Pock, "A first-order primal-dual algorithm for convex problems with applications to imaging," *J. Math. Imaging Vis.* **40**(1), 120–145 (2011).
27. H. Woo, and S. Yun, "Proximal linearized alternating direction method for multiplicative denoising," *SIAM J. Sci. Comput.* **35**(2), B336–B358 (2013).
28. S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers," *Found. Trends. Mach. Learn.* **3**(1), 1–122 (2011).
29. K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," *IEEE Trans. image process.* **16**(8), 2080–2095 (2007).
30. K. Dabov, A. Danieyan, and A. Foi, "BM3D MATLAB Software," Tampere University of Technology (2014) [retrieved 1 May 2017], <http://www.cs.tut.fi/~foi/GCF-BM3D/>.
31. M. Li, R. Idoughi, B. Choudhury, and W. Heidrich, "OCT Denoising Code," github (2017), <https://github.com/vccimaging/OCTDenoising>.
32. S. J. Chiu, X. T. Li, P. Nicholas, C. A. Toth, J. A. Izatt, S. Farsiu, "Automatic segmentation of seven retinal layers in SDOCT images congruent with expert manual segmentation," *Opt. Express* **18**(18), 19413–19428 (2010).

## 1. Introduction

Optical Coherence Tomography (OCT) is a coherence-gated imaging modality with vast applications in biomedicine because of its non-invasive nature and ability to image micro-scale 3D features. The visualization quality of OCT is often degraded by speckle noise as a natural consequence of limited light bandwidth, when there are multiple scatterers lying within the coherence length. The Rayleigh distribution is the first statistical model used to describe such speckle patterns [1], but it assumes perfect polarization, which in practice does not hold for real OCT systems. Recent investigations of speckle statistics consist of fitting probability distributions to the signal intensities within tissue. The Generalized Gamma distribution is found to be one of the best probabilistic models [2, 3]. The sensitivity of the spatial distribution of speckle to morphological changes in skin tissue has been used as a criterion for segmentating OCT images [4]. The goal of this work is to demonstrate that, when combined with suitable image regularization, statistical properties of speckles can be effectively exploited to produce cleaner visualization of OCT images.

First we summarize existing OCT speckle reduction techniques. According to the number of frames involved, OCT denoising methods can be grouped into two categories: single-frame denoising and multi-frame denoising. The most common multi-frame technique is to acquire uncorrelated OCT images of the same spot at multiple backscattering angles [5], or at different spatial positions [6]. An average is taken over all uncorrelated frames to improve the signal-to-noise ratio (SNR). By combining digital filtering and angular compounding, Ozcan et al. [7] demonstrate the potential for reducing the number of angles while achieving same speckle reduction results. However, this technique often requires hardware modification and complicated

acquisition processes. Apart from hardware modifications, there are also algorithmic approaches. Mayer et al. [8] introduce a wavelet-based method that utilizes 2D wavelet decomposition of single frames and re-weights the wavelet coefficients for reconstruction. This method requires acquisition of multiple frames throughout the whole investigation region, therefore leading to slower scanning speed. Jian et al. [9] improve the sparse representation of curved edges in OCT images using 3D curvelet transformation, enabling noise suppression of volume data without multiple scanning of the same region. Bian et al. [10] develop an optimization model combining an interframe low-rank prior with an intraframe anisotropic total variation (TV) prior. Fang et al. [11] propose a method that learns a sparse representation dictionary from a single averaged B-scan at a higher SNR and then use the learned features to denoise new low-SNR B-scans. The performance of this method relies on the quality of the high-SNR frame. Correlated noise, e.g. stationary speckle, can be recognized as features in the averaged frame, which then get introduced into other scans. Fang et al. [12] later introduce a single-frame version that does not require a priori high-quality frame, instead utilizing a previously collected datasets. A segmentation step is further included to form a segmentation based sparse reconstruction (SSR) framework which is specifically designed for retinal OCT images [13].

Single-frame approaches can be further divided into two groups: image-domain methods and wavelet-domain methods. Image-domain methods often adopt regularizers from the field of image processing, e.g. TV regularization assuming a Gamma distribution for the speckle [14], second-order total generalized variation (TGV) based on the classical Vese–Osher (VO) decomposition model [15, 16], and PDE-Based nonlinear diffusion [17]. Probabilistic approaches have also been developed for denoising OCT images. A conditional posterior sampling approach is introduced to obtain a Bayesian least squares estimation of noise-free images in the logarithm space with a stationary speckle variance [18]. A similar Bayesian approach assuming non-stationary speckle characteristics is further developed [19]. Recently the non-local means denoising filter has been adapted for OCT images and outperforms many other algorithms [20]. However, it performs well only when certain features appear repeatedly within the search window and suffers from a high computational cost. Wavelet-based methods exploit basic properties of wavelet coefficients of OCT images. Gaussian Scale Mixtures with Bayesian least square estimation [21], interval type II fuzzy based thresholding filtering [22], block-matching 3D (BM3D) based technique in the logarithm space [23], and dictionary-learning using complex wavelet based K-SVD [24] are recent methods developed in this area. Among these, the more recent BM3D and K-SVD methods tend to have the best performance in practice. The major drawback of wavelet-based approaches is the introduction of artifacts by operations in the wavelet domain. Most of the above methods assume a multiplicative noise model and thus take the logarithm of the measured data to convert the multiplicative relationship into an additive relationship.

As compared to aforementioned methods, we present a new approach that combines empirical speckle statistics with the Huber variant of total variation regularization that effectively remove speckle noise while simultaneously preserving edges. Our method makes the following contributions:

- While some existing methods assume a non-stationary speckle model, our statistical study shows that the speckle characteristics are well described by a stationary model based on two findings: first, the speckle is found to be multiplicative noise with a standard deviation proportional to the local mean intensity. Second, the probability distribution function for each mean intensity level can be converted into a Gaussian distribution with a square-root transformation.
- A novel optimization formulation based on our proposed speckle distribution with the Huber variant of TV regularization is proposed for speckle reduction of OCT images. Traditional TV regularization suffers from the so-called staircase artifacts, i.e. artificial flat

areas. The Huber penalty function acts as a remedy for this unwanted effect by modifying the  $L^1$  norm in the total variation term.

- A fast iterative algorithm for solving the proposed optimization problem is developed based on a linearized version of alternating direction method of multipliers (ADMM).

## 2. Methodology

### 2.1. Empirical statistical analysis

In the following, we empirically study statistical properties of OCT images that can later be used to derive a noise model. For this purpose, we work on two-dimensional raw OCT data obtained with two Thorlabs Ganymed SD-OCT devices with different bandwidths. The standard OCT reconstruction method is used to obtain a collection of 2D non-averaged B-scans by directly taking a Fourier transform on the frequency-domain measurement followed by taking absolute values. As the noise statistics is sensitive to any transformation, no post-processing is applied to the intensity image. The noise obtained from our devices follows a similar distribution to the Monte Carlo speckle statistics generated in [2] as well as their experimental results from a swept-source OCT system, which supports the applicability of our model to other types of SD-OCT devices. We further demonstrate the generality of our model by showing results on a publicly available dataset of retinal OCT images [25] in Sec. 3.4. Figure 1 shows two datasets from the first system. The first one (phantom) corresponds to a 3D-printed structure with alternating white and transparent layers (Fig. 1(a)). The second dataset shows a biofilm grown on a membrane inside water (Fig. 1(c)).

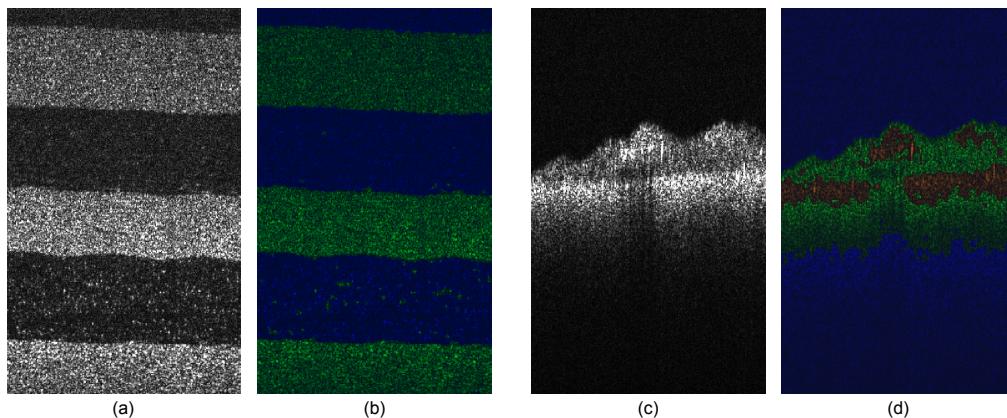


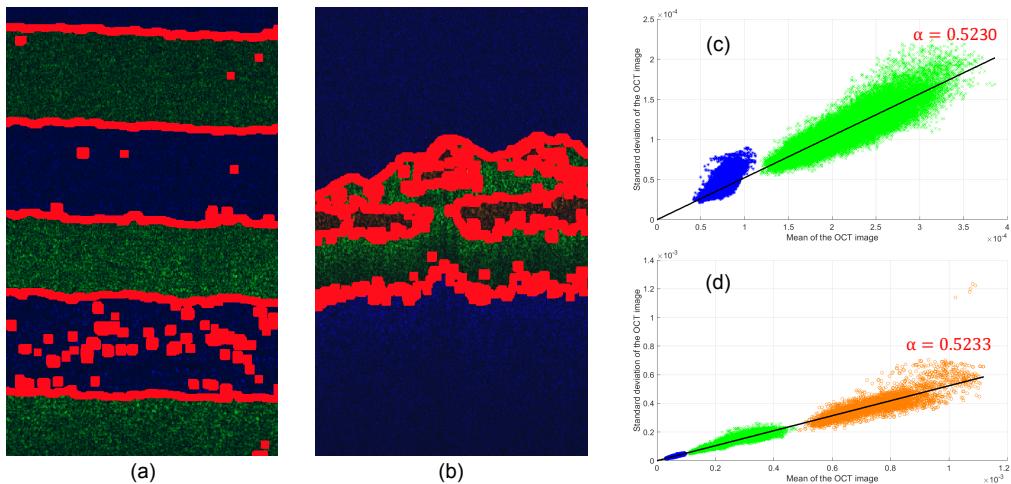
Fig. 1. Segmentation obtained for OCT images. (a) Original OCT image of phantom structure. (b) Segmented image of phantom structure, with 2 regions. (c) Original OCT image of a biofilm sample. (d) Segmented image of biofilm sample, with 3 regions.

We first study the local statistics of OCT image noise. To this end, we segment the image into regions of approximately homogeneous intensity, and compute the local mean and standard deviation for all pixels whose neighbors belong to the same cluster.

Figure 1 shows the obtained segmentations for the OCT image of the phantom (Fig. 1(b)) and biofilm, respectively (Fig. 1(d)). The 3D-printed phantom is composed of two different materials. This is well highlighted in the segmentation, except for some small regions, which in fact correspond to air bubbles. The biofilm image is composed of three relatively homogeneous regions: the background (blue), membrane (orange) and biofilm (green).

In Fig. 2, we illustrate the relationship between the local mean and the local standard deviation computed over the two images. These statistics were computed at each pixel location over a 9-by-9 window. In Fig. 2(a,b), the neighborhood of red pixels extends over two different clusters. Those pixels are not used for computing local statistics.

Figure 2(c,d) show that over the two images the local standard deviation is linearly related to the local mean. Moreover, the leading coefficient  $\alpha$  is the same for all clusters of the two images, and it is approximately equal to 0.523. One can notice that for the phantom image the variability around the fit is larger than for the biofilm image. This is due to the presence of a multitude of air bubbles introduced into the structure during the 3D-printing. These inclusions are not resolved outside the depth of field of the device.



**Fig. 2.** Illustration of the relationship between the local standard deviation and the local mean in OCT images. (a,b) Masks used for this computation for the images of phantom structure and biofilm sample respectively. For each pixel, the mean and standard deviation are computed over a local 9-by-9 window. Pixels lying between 2 different clusters (represented in red) are not considered in this computation. (c,d) Graphics showing the local standard deviation against the local mean, respectively for the phantom image and biofilm image.

This data shows also that the local statistics are spatially independent (stationary), since for two different windows having the same mean, the local standard deviation is the same regardless of location in the image.

Next, we study the noise distribution in homogeneous regions of the OCT images. Figure 3(a) shows an example of a homogeneous region selected from the 3D-printed phantom image. For those regions, the empirical probability density function (pdf) of the intensity values is computed (blue curve in Fig. 3(b)).

By applying a square-root transformation on the intensity values, we notice that the transformed pdf follows a Gaussian-like distribution (red curve in Fig. 3(b)). The black curve in Fig. 3(b) corresponds to a Gaussian fit of the square-root of the intensity values pdf. The similarity between the transformed pdf and its Gaussian fit is measured using a Q-Q plot, as shown in Fig. 3(c). The data points in the Q-Q plot lies approximately on the line  $y = x$ , which implies after the square-root transformation the intensity values in homogeneous regions follow a Gaussian distribution.

The statistical analysis presented above reveals three interesting properties of the speckle noise present in OCT images: (1) This noise is stationary (spatially independent) over the image, (2)

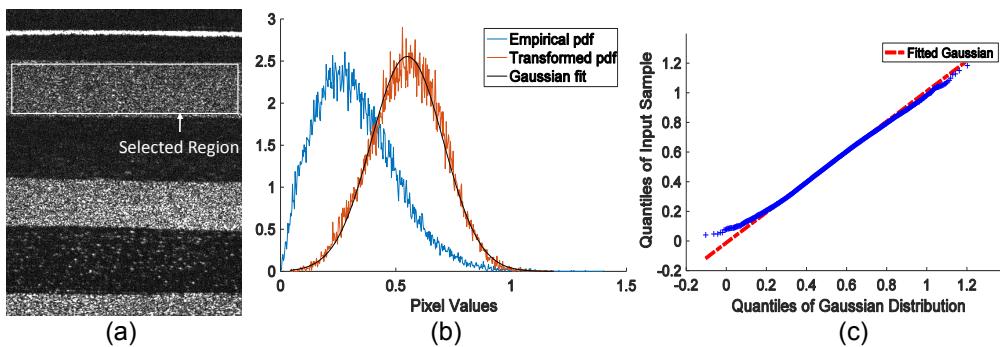


Fig. 3. (a) A selected homogeneous region of a 3D-printed phantom sample with layered structure. (b) Empirical probability distribution functions of intensity values in the selected region before (blue) and after (red) a square-root transformation, and the fitted Gaussian distribution (black) to the transformed distribution. (c) The Q-Q plot of the transformed distribution and the fitted Gaussian distribution. The dashed red line corresponds to quantiles of the fitted Gaussian distribution.

the standard deviation of the noise is linearly proportional to its mean, and (3) after a square-root transformation, the noise follows a Gaussian distribution. Moreover, the leading coefficient  $\alpha$  depends only on the device. It is constant for different datasets obtained with the same device but takes on different values for different devices.

In the next section we will introduce a new noise model, that takes into account these three properties.

## 2.2. Noise modeling

Next, we propose a stationary noise model that is sufficient to describe observations on speckle statistics from the previous section. Let  $\Omega$  denote the spatial domain of images. Let the image value at location  $\mathbf{x} \in \Omega$  be modeled as a random variable  $z$  given by

$$z(\mathbf{x}) = u(\mathbf{x}) \cdot s(\mathbf{x}), \quad (1)$$

where  $u(\mathbf{x})$  is the deterministic latent pixel value at  $\mathbf{x}$ , and the noise generator  $s(\mathbf{x})$  is a spatially-independent random variable with an unknown distribution and expected value  $\mathbb{E}(s) = 1$ . The three observations from above thus can be written as

$$\mathbb{E}(z(\mathbf{x})) = u(\mathbf{x}). \quad (2)$$

$$\sigma(z(\mathbf{x})) = \alpha u(\mathbf{x}) = \alpha \mathbb{E}(z(\mathbf{x})), \text{ and} \quad (3)$$

$$\sqrt{z(\mathbf{x})} \sim \mathcal{N}(\mu(\mathbf{x}), v(\mathbf{x})), \quad (4)$$

where  $\alpha$  is a constant relating the expectation of  $z(\mathbf{x})$  to its standard deviation, while  $\mathcal{N}(\mu(\mathbf{x}), v(\mathbf{x}))$  represents a Gaussian distribution with mean  $\mu(\mathbf{x})$  and variance  $v(\mathbf{x})$ .

The proposed noise model shares a similar form with the Generalized Gamma distribution as well as a common property [4], i.e. a linear relation between the mean and standard deviation of a random sample.

## 2.3. Model parameters

To justify the noise model proposed in (1) we now determine the parameters for the noise generator  $s$  that are consistent with our statistical observations listed in Eqs. (2)-(4).

We start by deriving the standard deviation  $\sigma$ . From Eqs. (1) we have the following relation between the standard deviation of  $z$  and that of  $s$ :  $\sigma(z(\mathbf{x}))^2 = u(\mathbf{x})^2\sigma(s)^2$ , i.e.  $\sigma(z(\mathbf{x})) = u(\mathbf{x})\sigma(s)$ . By comparison with Eq. (3), we obtain the relationship

$$\sigma(s) = \alpha. \quad (5)$$

Now we consider the noise distribution (4):  $\sqrt{z(\mathbf{x})} = \sqrt{u(\mathbf{x})}\sqrt{s(\mathbf{x})} \sim \mathcal{N}(\mu(\mathbf{x}), v(\mathbf{x}))$ , from which we obtain

$$\sqrt{s(\mathbf{x})} \sim \mathcal{N}\left(\frac{\mu(\mathbf{x})}{\sqrt{u(\mathbf{x})}}, \frac{v(\mathbf{x})}{u(\mathbf{x})}\right), \quad (6)$$

where  $\mathcal{N}\left(\frac{\mu(\mathbf{x})}{\sqrt{u(\mathbf{x})}}, \frac{v(\mathbf{x})}{u(\mathbf{x})}\right)$  denotes a Gaussian distribution with mean  $\frac{\mu(\mathbf{x})}{\sqrt{u(\mathbf{x})}}$  and variance  $\frac{v(\mathbf{x})}{u(\mathbf{x})}$ . The above expression for  $\sqrt{s(\mathbf{x})}$  is equivalent to

$$s(\mathbf{x}) \sim X^2 \quad \text{with } X \sim \mathcal{N}\left(\frac{\mu(\mathbf{x})}{\sqrt{u(\mathbf{x})}}, \frac{v(\mathbf{x})}{u(\mathbf{x})}\right), \quad (7)$$

Since  $s(\mathbf{x})$  is spatially-invariant, we also have

$$\frac{\mu(\mathbf{x})}{\sqrt{u(\mathbf{x})}} = c_1, \quad \frac{v(\mathbf{x})}{u(\mathbf{x})} = c_2 \quad \forall \mathbf{x} \in \Omega, \quad (8)$$

where  $c_1$  and  $c_2$  are two constants that can be derived by considering the moments of  $s \sim X^2$ :

$$\begin{aligned} \mathbb{E}(X^2) &= \left[ \frac{\mu(\mathbf{x})}{\sqrt{u(\mathbf{x})}} \right]^2 + \frac{v(\mathbf{x})}{u(\mathbf{x})} = \frac{\mu(\mathbf{x})^2 + v(\mathbf{x})}{u(\mathbf{x})}, \\ \mathbb{E}(X^4) &= \left[ \frac{\mu(\mathbf{x})}{\sqrt{u(\mathbf{x})}} \right]^4 + 6 \left[ \frac{\mu(\mathbf{x})}{\sqrt{u(\mathbf{x})}} \right]^2 \cdot \frac{v(\mathbf{x})}{u(\mathbf{x})} + 3 \left[ \frac{v(\mathbf{x})}{u(\mathbf{x})} \right]^2 = \frac{\mu(\mathbf{x})^4 + 6\mu(\mathbf{x})^2v(\mathbf{x}) + 3v(\mathbf{x})^2}{u(\mathbf{x})^2}. \end{aligned} \quad (9)$$

Therefore

$$1 = \mathbb{E}(s) = \mathbb{E}(X^2) = \frac{\mu(\mathbf{x})^2 + v(\mathbf{x})}{u(\mathbf{x})},$$

which means

$$\mu(\mathbf{x})^2 + v(\mathbf{x}) = u(\mathbf{x}). \quad (10)$$

Also

$$\alpha^2 = \sigma(s)^2 = \sigma(X^2)^2 = \mathbb{E}(X^4) - \mathbb{E}(X^2)^2,$$

which leads to

$$\alpha^2 = \frac{4\mu(\mathbf{x})^2v(\mathbf{x}) + 2v(\mathbf{x})^2}{u(\mathbf{x})^2}. \quad (11)$$

Combining (10) and (11) we obtain

$$c_1 = \frac{\mu(\mathbf{x})}{\sqrt{u(\mathbf{x})}} = \left(1 - \frac{\alpha^2}{2}\right)^{\frac{1}{4}} \quad \text{and} \quad c_2 = \frac{v(\mathbf{x})}{u(\mathbf{x})} = 1 - \left(1 - \frac{\alpha^2}{2}\right)^{\frac{1}{2}}. \quad (12)$$

## 2.4. Optimization problem for noise removal

To turn this model into a practical noise removal algorithm, we now treat  $u$  as a random variable and derive the maximum-a-posteriori estimate of  $u$ . A prior probability distribution  $g(u)$  is introduced to express prior knowledge about the latent image, e.g.  $g(u) \propto \exp(-\lambda \int_{\Omega} |\nabla u|)$  representing the total variation prior where  $\nabla$  is the gradient operator. According to Bayes' rule, the likelihood  $f(u|z)$  of the latent image  $u$  given the noisy observation  $z$  satisfies  $f(u|z) \propto g(u)f(z|u)$ , where  $f(z|u)$  can be directly derived from the noise model (1).

Let  $\mathbf{z}, \mathbf{u}$  denote the discretized noisy and latent images.  $\frac{\mathbf{z}_{i,j}}{\mathbf{u}_{i,j}}$  follows the distribution of a squared Gaussian random variable, i.e.  $\frac{\mathbf{z}_{i,j}}{\mathbf{u}_{i,j}} \sim X^2$  where  $X$  is a Gaussian random variable for an arbitrary pixel location  $(i, j)$ . Applying a transformation of variables and the chain rule, we get the pdf of  $\mathbf{z}|\mathbf{u}$

$$f(\mathbf{z}|\mathbf{u}) = \prod_{i,j} \frac{1}{\sqrt{2c_2\pi}} \exp\left(-\frac{(\sqrt{\frac{\mathbf{z}_{i,j}}{\mathbf{u}_{i,j}}} - c_1)^2}{2c_2}\right) \frac{1}{2\sqrt{\mathbf{z}_{i,j}\mathbf{u}_{i,j}}} . \quad (13)$$

The maximum-a-posteriori estimate  $\mathbf{u}_{\text{MAP}}$  is given by

$$\begin{aligned} \mathbf{u}_{\text{MAP}} &= \underset{\mathbf{u} \in \mathbf{U}}{\operatorname{argmax}} g(\mathbf{u}) f(\mathbf{z}|\mathbf{u}) \\ &= \underset{\mathbf{u} \in \mathbf{U}}{\operatorname{argmax}} \log(g(\mathbf{u})) + \log(f(\mathbf{z}|\mathbf{u})) \\ &= \underset{\mathbf{u} \in \mathbf{U}}{\operatorname{argmin}} -\log(g(\mathbf{u})) + \frac{1}{2c_2} \left\| \sqrt{\frac{\mathbf{z}}{\mathbf{u}}} - c_1 \right\|_2^2 + \frac{1}{2} \sum_{i,j} \log(\mathbf{u}_{i,j}) . \end{aligned} \quad (14)$$

where  $\mathbf{U}$  is the solution space. However, the model (14) is highly non-convex, hence it is difficult to find a global solution. We suggest a log transform to improve convexity. Define  $\hat{\mathbf{U}} = \log \mathbf{U}$  as the log-transformed image space. The variational model for the transformed image can be formulated as follows:

$$\min_{\mathbf{u} \in \hat{\mathbf{U}}} -\log(g(\mathbf{u})) + \frac{1}{2c_2} \left\| \sqrt{\mathbf{z}} e^{-\frac{1}{2}\mathbf{u}} - c_1 \right\|_2^2 + \frac{1}{2} \sum_{i,j} \mathbf{u}_{i,j} . \quad (15)$$

This new model is called the exponential model. Since the logarithm is a monotonically increasing function, the location of edges does not change, and thus edge-preserving priors can be applied to the log-transformed image instead of the original image.

Specifically, we propose  $g(\mathbf{u}) \propto \exp(-\lambda \|\nabla \mathbf{u}\|_H)$ , where  $\lambda$  is a weight parameter on the edge-preserving prior and  $\|\nabla \mathbf{u}\|_H$  represents the Huber variant of the total variation regularization, and is given by

$$\|\nabla \mathbf{u}\|_H = \sum_{i,j} |(\nabla \mathbf{u})_{i,j}|_\beta , \quad (16)$$

where  $(\nabla \mathbf{u})_{i,j}$  denotes the gradient of  $\mathbf{u}$  at pixel  $(i, j)$ , and  $|\cdot|_\beta$  denotes the Huber penalty function for vectors in  $\mathbb{R}^2$ :

$$|(\nabla \mathbf{u})_{i,j}|_\beta = \begin{cases} \frac{|(\nabla \mathbf{u})_{i,j}|}{2\beta} & \text{if } |(\nabla \mathbf{u})_{i,j}| \leq \beta \\ |(\nabla \mathbf{u})_{i,j}| - \frac{\beta}{2} & \text{if } |(\nabla \mathbf{u})_{i,j}| > \beta \end{cases}$$

where  $\beta > 0$  is a small parameter defining the transition from quadratic regularization to total variation regularization. While the standard TV prior exhibits strong staircase artifacts in the reconstruction, the Huber variant leads to more natural results [26]. The quadratic regularization for low gradient values leads to piecewise smoothness in regions with gradient magnitude smaller than  $\beta$ . The parameter  $\beta$  offers an extra degree of freedom in the shape of the penalty function. When  $\beta$  is set to 0,  $\|\nabla \mathbf{u}\|_H$  coincides with the standard TV regularization.

## 2.5. Noise removal algorithm

A proximal linearized alternating direction (PLAD) algorithm [27] can be employed to solve the optimization problem (15). Denote the data fidelity term as  $f(\mathbf{u}) = \frac{1}{2c_2} \left\| \sqrt{\mathbf{z}} e^{-\frac{1}{2}\mathbf{u}} - c_1 \right\|_2^2 +$

$\frac{1}{2} \sum_{i,j} \mathbf{u}_{i,j}$ . The augmented Lagrangian associated with (15) is given by

$$L_\gamma(\mathbf{u}, \mathbf{w}, \mathbf{p}) := f(\mathbf{u}) + \lambda \|\mathbf{w}\|_H + \langle \mathbf{p}, \mathbf{w} - \nabla \mathbf{u} \rangle + \frac{\gamma}{2} \|\mathbf{w} - \nabla \mathbf{u}\|_2^2. \quad (17)$$

Directly applying the well-known ADMM [28] leads to solving a nonlinear system when updating  $\mathbf{u}$  in each iteration, without a closed-form solution. A Taylor expansion is proposed in [27] for the smooth function  $f(\mathbf{u})$  and the quadratic function  $\frac{\gamma}{2} \|\mathbf{w} - \nabla \mathbf{u}\|_2^2$ :

$$f(\mathbf{u}) + \frac{\gamma}{2} \|\mathbf{w} - \nabla \mathbf{u}\|_2^2 \approx f(\mathbf{u}_0) + \frac{\gamma}{2} \|\mathbf{w} - \nabla \mathbf{u}_0\|_2^2 + \langle \nabla_{\mathbf{u}} f(\mathbf{u}_0) + \gamma \operatorname{div}(\mathbf{w} - \nabla \mathbf{u}_0), \mathbf{u} - \mathbf{u}_0 \rangle + \frac{1}{2\delta} \|\mathbf{u} - \mathbf{u}_0\|_2^2. \quad (18)$$

Therefore for each iteration of  $\mathbf{u}$  update we replace the augmented Lagrangian in Eq. (17) with the approximation

$$\begin{aligned} L_\gamma(\mathbf{u}, \mathbf{w}, \mathbf{p}) := & f(\mathbf{u}^k) + \frac{\gamma}{2} \|\mathbf{w} - \nabla \mathbf{u}^k\|_2^2 + \langle \nabla_{\mathbf{u}} f(\mathbf{u}^k) + \gamma \operatorname{div}(\mathbf{w} - \nabla \mathbf{u}^k), \mathbf{u} - \mathbf{u}^k \rangle + \frac{1}{2\delta} \|\mathbf{u} - \mathbf{u}^k\|_2^2 \\ & + \lambda \|\mathbf{w}\|_H + \langle \mathbf{p}, \mathbf{w} - \nabla \mathbf{u} \rangle, \end{aligned} \quad (19)$$

where  $\mathbf{u}^k$  is the value of  $\mathbf{u}$  from the previous iteration. The final method is summarized in Algorithm 1.

---

**Algorithm 1** Our denoising algorithm

---

**Input:** the noisy observation  $\mathbf{z}$ , constant parameter  $c_1, c_2$ , weight  $\lambda$ , Huber parameter  $\beta$ ; Given the initialization  $\mathbf{u}^0, \mathbf{w}^0, \mathbf{p}^0$ . Choose  $\gamma > 0$  and  $\delta > 0$ . Implement the following iteration until convergence:  
1:  $\mathbf{u}^{k+1} = \underset{\mathbf{u} \in \hat{\mathcal{U}}}{\operatorname{argmin}} \langle \nabla_{\mathbf{u}} f(\mathbf{u}^k) + \gamma \operatorname{div}(\mathbf{w}^k - \nabla \mathbf{u}^k), \mathbf{u} - \mathbf{u}^k \rangle + \langle \mathbf{p}^k, \mathbf{w}^k - \nabla \mathbf{u} \rangle + \frac{1}{2\delta} \|\mathbf{u} - \mathbf{u}^k\|_2^2;$   
2:  $\mathbf{w}^{k+1} = \underset{\mathbf{w}}{\operatorname{argmin}} \lambda \|\mathbf{w}\|_H + \langle \mathbf{p}^k, \mathbf{w} - \nabla \mathbf{u}^{k+1} \rangle + \frac{\gamma}{2} \|\mathbf{w} - \nabla \mathbf{u}^{k+1}\|_2^2;$   
3:  $\mathbf{p}^{k+1} = \mathbf{p}^k + \gamma(\mathbf{w}^{k+1} - \nabla \mathbf{u}^{k+1}).$

**Output:**  $\mathbf{u} \leftarrow \mathbf{u}^{k+1}$ , the estimated latent image.

---

Step 1 and Step 2 in the above iteration of Algorithm 1 can be computed using the following simplified form:

$$\begin{aligned} 1: \mathbf{u}^{k+1} &= \mathbf{u}^k - \delta \left( \frac{1}{2} - \frac{\sqrt{\mathbf{z}}}{2c_2} e^{-\frac{1}{2}\mathbf{u}} (\sqrt{\mathbf{z}} e^{-\frac{1}{2}\mathbf{u}} - c_1) + \gamma \operatorname{div}(\mathbf{w}^k - \nabla \mathbf{u}^k) + \operatorname{div}(\mathbf{p}^k) \right), \\ 2: \mathbf{w}_{i,j}^{k+1} &= \max \left( \left\| (\nabla \mathbf{u}^{k+1})_{i,j} - \frac{\mathbf{p}_{i,j}^k}{\gamma} \right\|_2 - \frac{\lambda}{\gamma}, \frac{\left\| (\nabla \mathbf{u}^{k+1})_{i,j} - \frac{\mathbf{p}_{i,j}^k}{\gamma} \right\|_2}{1 + \frac{\lambda}{\gamma\beta}} \right) \frac{(\nabla \mathbf{u}^{k+1})_{i,j} - \frac{\mathbf{p}_{i,j}^k}{\gamma}}{\left\| (\nabla \mathbf{u}^{k+1})_{i,j} - \frac{\mathbf{p}_{i,j}^k}{\gamma} \right\|_2}, \end{aligned} \quad (20)$$

where  $(\nabla \mathbf{u}^{k+1})_{i,j} = ((\partial_x \mathbf{u}^{k+1})_{i,j}, (\partial_y \mathbf{u}^{k+1})_{i,j}) \in \mathbb{R}^2$  denote the gradient of  $\mathbf{u}^{k+1}$  at pixel  $(i, j)$  and  $\mathbf{p}_{i,j}^k \in \mathbb{R}^2$ .

### 3. Results

To evaluate the merits of the proposed method, we compare against 5 state-of-the-art speckle reduction methods: multiscale sparsity based tomographic denoising (MSBTD) with the log-space transformation [11], log-space BM3D [29], complex wavelet based K-SVD [24], general Bayesian estimation method [18], and TGV decomposition [16]. All processing is implemented

in MATLAB R2016b on a computer with an Intel Xeon E5-2680 v3 CPU. Four types of samples are used in the following for evaluating the performance of the proposed method: a 3D-printed structure with alternating layers of two different materials, a biofilm grown under water on a membrane, orange pulp, and finally the skin of a chicken breast. The original noisy images are shown in Fig. 5.

### 3.1. Parameter selection

This section presents a qualitative analysis of the parameters involved in the proposed method, and reports the parameters used in all compared methods.

For the proposed method, the parameters  $c_1$  and  $c_2$  are constants determined by the device-dependent parameter  $\alpha$ , i.e., the ratio between local standard deviation and local mean in segmented images as shown in Fig. 2(a,b). Therefore  $c_1$  and  $c_2$  can be pre-computed for each device. The two regularization parameters,  $\lambda$  on TV regularization and  $\beta$  on Huber penalty are chosen by the users. A higher value for  $\lambda$  leads to better noise suppression but tends to remove detailed structures, and larger  $\beta$  yields smoother transition between image gradients. In our experiments we choose  $\lambda \approx 0.4$  and  $\beta = 0.02$ . In general,  $\lambda$  should be proportional to the noise variance and  $\beta$  should be kept small.

The noise level is estimated on log-transformed images by averaging the standard deviation in homogeneous regions after segmentation, as shown in Fig. 2(a,b). All compared methods are implemented using parameters proposed in the associated literature, with our estimated noise level, if required, as an input.

### 3.2. OCT denoising results

The denoised images of a randomly selected B-scan from each sample dataset after applying the above denoising methods are depicted in Fig. 4. The grainy patterns of speckle degrades image quality by reducing contrast and making fine structure more difficult to resolve. It is easy to observe that both the MSBTD and BM3D approaches perform well in low-intensity areas but fail to reduce static speckle in high-intensity regions where grainy patterns are dominant (Fig. 4(1-b,c) and Fig. 4(4-b,c)). The K-SVD method suppresses to a good extent throughout the whole image but the results are contaminated by randomly scattered dark spots (Fig. 4(2-d) and Fig. 4(4-d)), thus leading to ambiguity interpreting fine details in certain occasions. The general Bayesian approach has a consistent but overall unsatisfactory performance across structural components. We notice that the TGV decomposition method provides improved speckle suppression performance in low-intensity regions. However, it removes detailed structure due to overcompensation (Fig. 4(3-f)). The proposed method achieves a balanced performance in speckle reduction and structure preservation. It also preserves a high contrast for structural details, e.g. air bubbles in the upper region in Fig. 4(1-g).

The quantitative comparison of the performance of these methods is based on the widely-used metrics such as contrast-to-noise ratio (CNR), equivalent number of looks (ENL), edge preservation (EP) as well as the runtime.

### 3.3. Quantitative analysis

In this section, quantitative metrics are calculated for the Regions of interest (ROIs) of each sample data and the results are shown in Table 1.

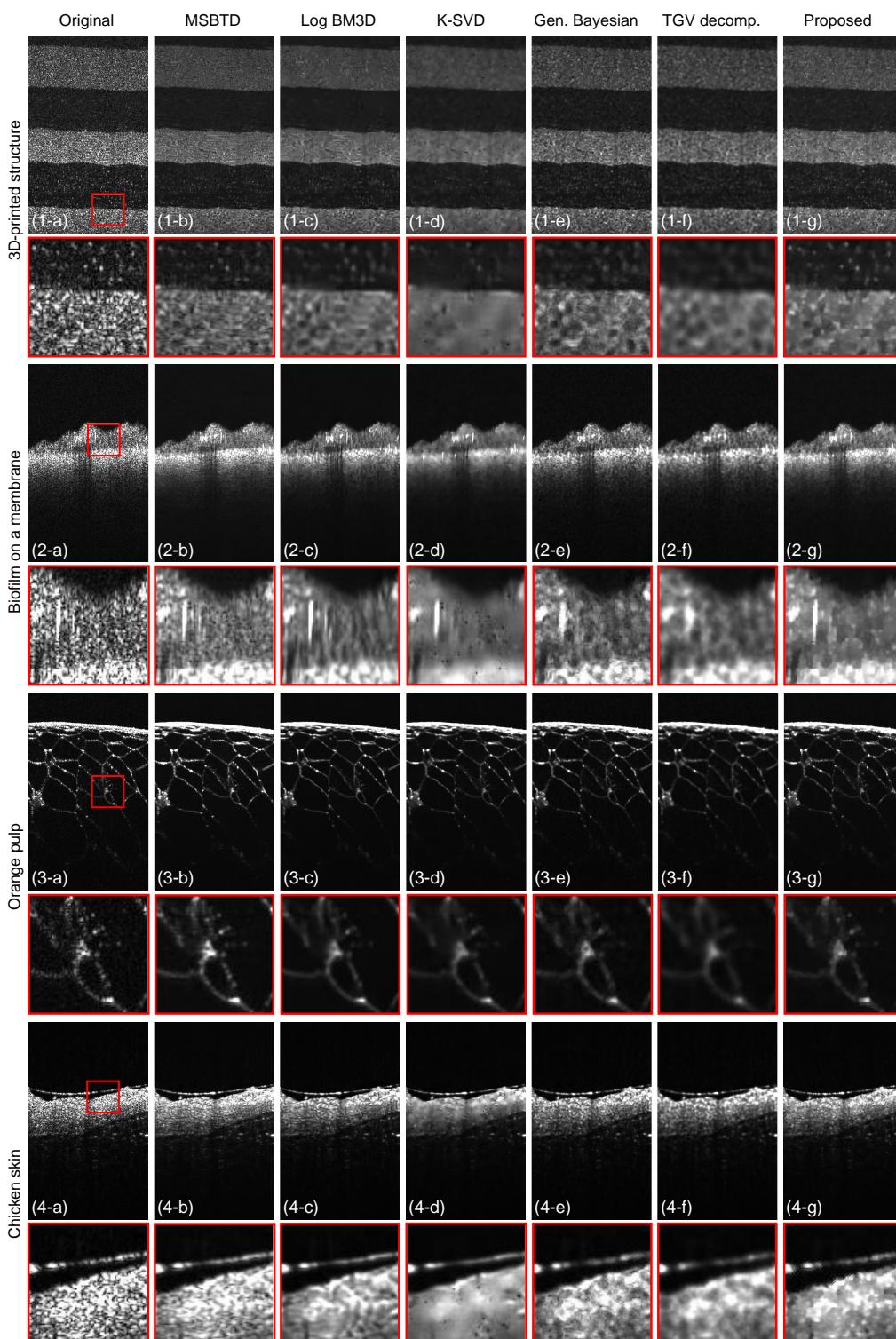


Fig. 4. Denoising results of 4 sample images. Odd rows: whole images. Even rows: zoomed-in images of the selected region.

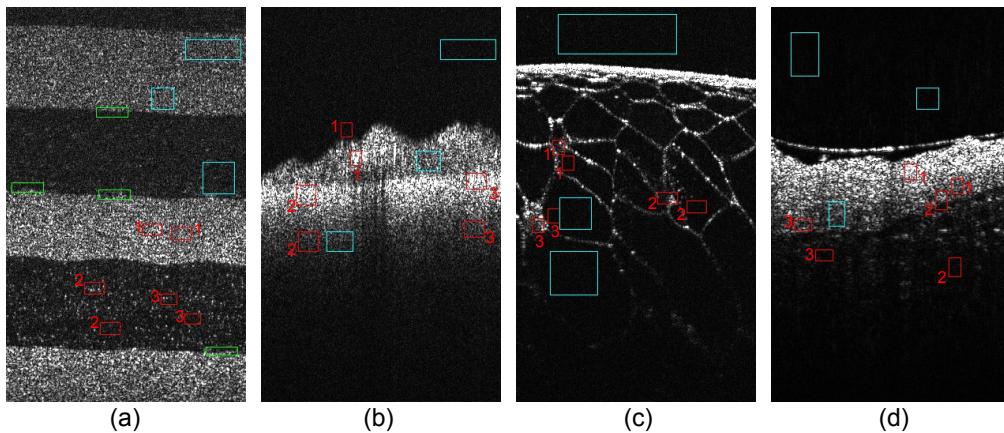


Fig. 5. Original noisy images with selected regions of interest (ROIs) marked in boxes for calculation of: CNR (red), ENL (cyan), EP (green). (a) 3D-printed phantom with layered structure. (b) Biofilm on a membrane. (c) Orange pulp. (d) Chicken skin.

Table 1. Performance metrics computed for different methods. The Average is taken over corresponding ROIs and over all samples. The standard deviation of each metric is also included. \* EP is calculated only on the 3D-printed phantom sample.

	Original	MSBTD	Log-space BM3D	K-SVD	General Bayesian	TGV decomp.	Proposed
CNR	$1.55 \pm 0.87$	$2.34 \pm 1.58$	$2.47 \pm 1.31$	$3.01 \pm 1.79$	$2.21 \pm 1.29$	$2.57 \pm 1.43$	<b>3.07 ± 1.88</b>
ENL	$3.53 \pm 0.20$	$28.01 \pm 16.80$	<b>157.55 ± 198.29</b>	$76.33 \pm 42.96$	$20.40 \pm 7.08$	$38.84 \pm 19.95$	$64.19 \pm 41.46$
EP*	$0.68 \pm 0.14$	<b>0.81 ± 0.01</b>	$0.79 \pm 0.01$	$0.80 \pm 0.01$	$0.78 \pm 0.01$	$0.77 \pm 0.01$	<b>0.81 ± 0.01</b>
Runtime (sec)	-	$394.10 \pm 8.39$	$2.91 \pm 0.13$	$11.96 \pm 0.42$	$39.29 \pm 0.56$	$10.62 \pm 0.45$	<b>2.25 ± 0.11</b>

### 3.3.1. Contrast-to-noise ratio (CNR)

CNR is a measure of the contrast between a feature in ROI and the noisy background. The ROIs for SNR evaluation are marked by red boxes in Fig. 5. The CNR over  $r$ -th ROI is defined as [11]

$$\text{CNR}_r = \frac{|\mu_r - \mu_b|}{\sqrt{0.5(\sigma_r^2 + \sigma_b^2)}} , \quad (21)$$

where  $\mu_r$  and  $\sigma_r^2$  denote the mean and variance of the  $r$ -th ROI.  $\mu_b$  and  $\sigma_b^2$  denote the mean and variance of the background reference region.

To take into account multiple ROIs, the average CNR over a number of  $N$  ROIs is computed by

$$\text{CNR} = \frac{1}{N} \sum_{r=1}^N \text{CNR}_r . \quad (22)$$

The values of CNR depend significantly on features of the selected region, leading to a large standard deviation across ROIs on different sample images. As shown in Table 1, the proposed

method is able to achieve the highest average CNR but also with the highest standard deviation.

### 3.3.2. Equivalent number of looks (ENL)

ENL acts as an indicator for smoothness in a homogeneous region. A higher ENL value indicates the noise is better reduced from the homogeneous region. The ROIs for ENL analysis are marked by cyan boxes in Fig. 5. The average ENL over  $N$  ROIs can be calculated using the following definition [19]

$$\text{ENL} = \frac{1}{N} \sum_{r=1}^N \frac{\mu_r^2}{\sigma_r^2}, \quad (23)$$

where  $\mu_r$  and  $\sigma_r$  denote the mean and standard deviation of the  $r$ -th homogeneous ROI.

It is obvious that, in Table 1, BM3D obtains much higher ENL than other methods. However, we note that in both cases the standard deviation exceeds the mean of the metric, which implies that the performance of the method relies heavily on the choice of ROIs. Looking closely at Fig. 4(1-c), it can be observed that BM3D creates unwanted artifacts in high-intensity homogeneous regions. This means the performance of BM3D is significantly affected by speckle patterns whilst the proposed method is capable of removing both static and dynamic speckle noise.

### 3.3.3. Edge preservation (EP)

EP is a performance measure that computes the local correlation between the edges in the denoised image and in a “ground truth” reference image. To obtain a noise-reduced reference image close to the ground truth, we introduced slight movement of the sample during data acquisition and take average over 200 B-scans. As a result, this suppresses the stationary speckle in the final averaged image. The ROIs selected for calculating EP are marked by green boxes in Fig. 5. We compute an average EP over  $N$  selected ROIs [19]:

$$\text{EP} = \frac{1}{N} \sum_{r=1}^N \frac{\sum_{i,j \in \text{ROI}_r} (\Delta \mathbf{i}_r - \overline{\Delta \mathbf{i}_r}) \cdot (\Delta \mathbf{u}_r - \overline{\Delta \mathbf{u}_r})}{\sqrt{\sum_{i,j \in \text{ROI}_r} (\Delta \mathbf{i}_r - \overline{\Delta \mathbf{i}_r})^2 \cdot \sum_{i,j \in \text{ROI}_r} (\Delta \mathbf{u}_r - \overline{\Delta \mathbf{u}_r})^2}}, \quad (24)$$

where  $\Delta$  is the Laplacian operator.  $\mathbf{i}_r$  and  $\mathbf{u}_r$  are sub-matrices that represent the  $r$ -th ROI in the reference and denoised images, respectively.  $\overline{\Delta \mathbf{i}_r}$  and  $\overline{\Delta \mathbf{u}_r}$  denotes the empirical mean of  $\Delta \mathbf{i}_r$  and  $\Delta \mathbf{u}_r$  over  $3 \times 3$  neighborhoods. The range of EP is between 0 and 1, with a value close to 1 implying high similarity in edges. The proposed method and MSBTD achieve the best edge preserving performance as compared to others, as indicated in Table 1.

### 3.3.4. Runtime

Another important metric to assess algorithm performance is the runtime. Images of different sizes (from  $300 \times 500$  to  $500 \times 1000$ ) are used for testing method efficiency. All methods except K-SVD have computational time proportional to the image size, whereas the runtime for K-SVD increases at a sublinear rate. Therefore, K-SVD has higher efficiency for sufficiently large images. For a normal-sized image, however, the proposed method and BM3D are less time-consuming. Table 1 records the runtime on test images of size  $304 \times 1000$ . As indicated, our algorithm is the fastest among all, despite being entirely implemented in Matlab, while the second fastest method, BM3D relied on a highly optimized implementation [30]. This suggests a significant potential for further performance improvements for our method, especially with the help of GPU implementations. Such performance improvements will be particularly interesting when extending the method to 3D data.

### 3.4. Visual analysis on retinal data

One important application of OCT is in the field of ophthalmology. Denoising retinal OCT images serves as an important tool for diagnosis. The proposed method can be effectively applied to retinal OCT data. We use the retinal OCT dataset provided by Mahmudi et al. [25]. The acquired images are first re-scaled to produce a proper empirical probability density function, before applying the proposed method. The noise level is estimated in the same way, as before, by averaging the standard deviation of log-space intensity in flat regions. An efficient way to evaluate the performance of denoising methods on retinal data is to check the segmentation quality on denoised images with a specialized retinal segmentation algorithm. We use an automatic graph search algorithm [32].

The segmentation results of images denoised by all compared methods except MSBTD is shown in Fig. 6, with parameters chosen as suggested in the associated literature and kept the same for all methods. MSBTD method requires one high-SNR image as input, which is not provided in the retinal dataset. Therefore MSBTD is excluded from this analysis. The segmentation algorithm by default uses a Gaussian filter as the denoising method. The segmentation obtained with the default approach is shown in Fig. 6(a). The Gaussian filter is replaced by the comparative methods of our evaluation in Fig. 6(b-f). Unlike BM3D and K-SVD as well as the default Gaussian filter, the general Bayesian approach, TGV decomposition and the proposed method yield similar high-quality segmentation results as depicted by the orange curves in the corresponding figures (Fig. 6(d,e,f)) which correctly separate the bottom layer. The lower performance of BM3D and K-SVD can be attributed to the unwanted artifacts generated in the region of interest.

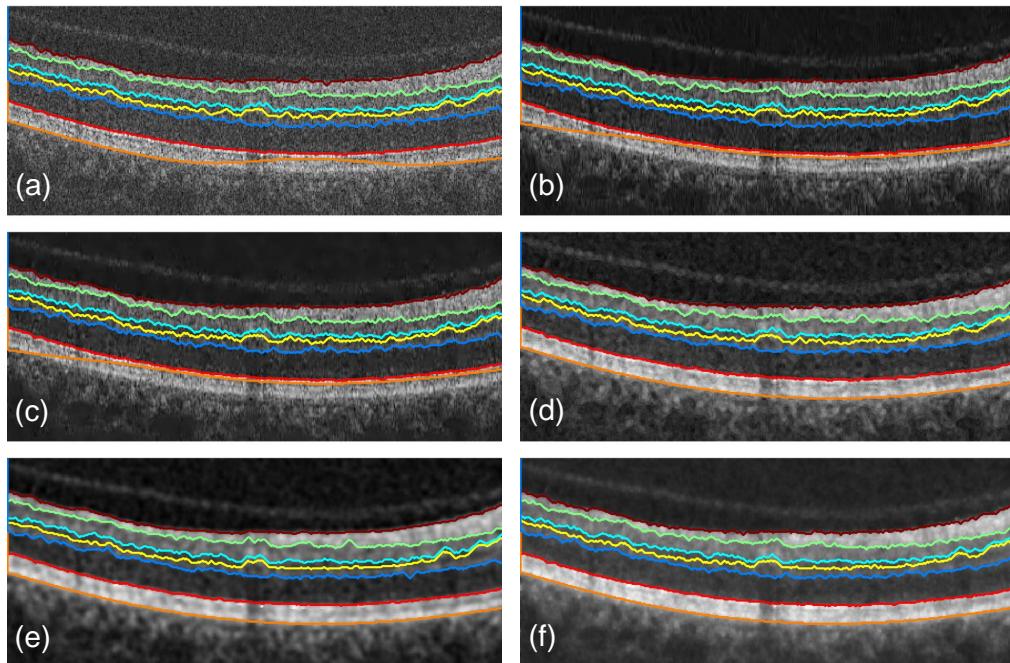


Fig. 6. Comparison of segmentation results on denoised retinal images using: (a) Gaussian filter. (b) Log-space BM3D. (c) K-SVD. (d) General Bayesian. (e) TGV decomposition. (f) Proposed.

#### 4. Conclusion

In this paper, we have introduced a new variational approach for denoising OCT images based on a new statistical model of spatially-invariant noise. The statistical analysis of empirical data demonstrates that the proposed probability distribution provides a good fit to the experimental results. It is shown in terms of CNR, ENL and EP that the proposed method achieves state-of-the-art performance in speckle noise reduction as well as edge preservation. A qualitative assessment shows that the MSBTD, log-space BM3D, general Bayesian and TGV decomposition methods have limited speckle reduction ability as compared to K-SVD and the proposed method. The K-SVD method provides superior speckle reduction and edge preservation performance but introduces unwanted dark spots in large smooth areas, which is not the case for the proposed method. Visual analysis on segmenting retinal images shows that the proposed method offers among the best improvement of the segmentation performance.

Although the proposed method is only applied to 2D B-scans in the experiment, adaptation to 3D volume data or multiple frames of B-scans are algorithmically straightforward, by extending the Huber total variation regularization to the third dimension. The smoothness constraint on the third dimension or across time is in fact expected to improved denoising performance as extra information can be extracted from neighboring frames.

The Matlab source code for our method is publicly available as [Code 1](#) (Ref. [31]). In the future, we will further improve the performance of the method and extend it to allow for real-time 3D applications.

#### Funding

King Abdullah University of Science and Technology (KAUST) (Visual Computing Center Competitive Funding)

#### Acknowledgments

We would like to thank Prof. TorOve Leiknes and Luca Fortunato of Water Desalination and Reuse Center, KAUST for help with the OCT data acquisition. We also take this opportunity to thank the ThorLabs, Munich engineering team for their assistance with the native OCT file handling procedures. The authors would also like to thank Mohamed Aly for his valuable feedback on the project.

#### Disclosures

The authors declare that there are no conflicts of interest related to this article.