Sparsity based denoising of spectral domain optical coherence tomography images

Leyuan Fang,[1,2,*] Shutao Li,[1] Qing Nie,[4,2] Joseph A. Izatt,[3,2] Cynthia A. Toth,[2,3] and Sina Farsiu[2,3]

[1]College of Electrical and Information Engineering, Hunan University, Changsha, 410082, China
[2]Department of Ophthalmology, Duke University Medical Center, Durham, NC, 27710, USA
[3]Department of Biomedical Engineering, Duke University, Durham, NC, 27708, USA
[4]School of Information and Electronics, Beijing Institute and Technology, Beijing, 100081, China
*leyuan.fang@duke.edu

Abstract: In this paper, we make contact with the field of compressive sensing and present a development and generalization of tools and results for reconstructing irregularly sampled tomographic data. In particular, we focus on denoising Spectral-Domain Optical Coherence Tomography (SDOCT) volumetric data. We take advantage of customized scanning patterns, in which, a selected number of B-scans are imaged at higher signal-to-noise ratio (SNR). We learn a sparse representation dictionary for each of these high-SNR images, and utilize such dictionaries to denoise the low-SNR B-scans. We name this method multiscale sparsity based tomographic denoising (MSBTD). We show the qualitative and quantitative superiority of the MSBTD algorithm compared to popular denoising algorithms on images from normal and age-related macular degeneration eyes of a multi-center clinical trial. We have made the corresponding data set and software freely available online.

© 2012 Optical Society of America

OCIS codes: (110.4500) Optical coherence tomography; (170.4460) Ophthalmic optics and devices; (100.0100) Image processing; (030.4280) Noise in imaging systems; (100.2980) Image enhancement; (170.5755) Retina scanning.

References and links

1. M. Choma, M. Sarunic, C. Yang, and J. Izatt, "Sensitivity advantage of swept source and Fourier domain optical coherence tomography," Opt. Express **11**(18), 2183–2189 (2003).
2. D. Huang, E. A. Swanson, C. P. Lin, J. S. Schuman, W. G. Stinson, W. Chang, M. R. Hee, T. Flotte, K. Gregory, C. A. Puliafito, and J. G. Fujimoto, "Optical coherence tomography," Science **254**(5035), 1178–1181 (1991).
3. J. M. Schmitt, S. H. Xiang, and K. M. Yung, "Speckle in optical coherence tomography," J. Biomed. Opt. **4**(1), 95–105 (1999).
4. 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).
5. J. Portilla, V. Strela, M. J. Wainwright, and E. P. Simoncelli, "Adaptive Wiener denoising using a Gaussian scale mixture model in the wavelet domain," in *2001 International Conference on Image Processing, 2001. Proceedings* (2001), Vol. 2, pp. 37–40.
6. H. Takeda, S. Farsiu, and P. Milanfar, "Robust kernel regression for restoration and reconstruction of images from sparse noisy data," in *2006 IEEE International Conference on Image Processing* (2006), pp. 1257–1260.
7. S. G. Chang, B. Yu, and M. Vetterli, "Adaptive wavelet thresholding for image denoising and compression," IEEE Trans. Image Process. **9**(9), 1532–1546 (2000).
8. P. Chatterjee and P. Milanfar, "Practical bounds on image denoising: from estimation to information," IEEE Trans. Image Process. **20**(5), 1221–1233 (2011).
9. 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).
10. M. Gargesha, M. W. Jenkins, A. M. Rollins, and D. L. Wilson, "Denoising and 4D visualization of OCT images," Opt. Express **16**(16), 12313–12333 (2008).
11. 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," Opt. Lett. **29**(24), 2878–2880 (2004).
12. 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," Opt. Express **18**(2), 1024–1032 (2010).

13. A. W. Scott, S. Farsiu, L. B. Enyedi, D. K. Wallace, and C. A. Toth, "Imaging the infant retina with a hand-held spectral-domain optical coherence tomography device," Am. J. Ophthalmol. **147**(2), 364–373.e2 (2009).
14. 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).
15. E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," IEEE Trans. Inf. Theory **52**(2), 489–509 (2006).
16. D. L. Donoho, "Compressed sensing," IEEE Trans. Inf. Theory **52**(4), 1289–1306 (2006).
17. D. Healy and D. J. Brady, "Compression at the physical interface," IEEE Signal Process. Mag. **25**(2), 67–71 (2008).
18. S. Ji, Y. Xue, and L. Carin, "Bayesian compressive sensing," IEEE Trans. Signal Process. **56**(6), 2346–2356 (2008).
19. S. J. Wright, R. D. Nowak, and M. A. T. Figueiredo, "Sparse reconstruction by separable approximation," IEEE J. Sel. Top. Signal Process. **57**, 2479–2493 (2009).
20. M. A. Neifeld and J. Ke, "Optical architectures for compressive imaging," Appl. Opt. **46**(22), 5293–5303 (2007).
21. R. M. Willett, R. F. Marcia, and J. M. Nichols, "Compressed sensing for practical optical imaging systems: a tutorial," Opt. Eng. **50**(7), 072601–072613 (2011).
22. "Age-Related Eye Disease Study 2 Ancillary Spectral Domain Optical Coherence Tomography Study (A2ASDOCT)" (2008–2013), *http://clinicaltrials.gov/ct2/show/NCT00734487*
23. S. Farsiu, J. Christofferson, B. Eriksson, P. Milanfar, B. Friedlander, A. Shakouri, and R. Nowak, "Statistical detection and imaging of objects hidden in turbid media using ballistic photons," Appl. Opt. **46**(23), 5805–5822 (2007).
24. X. Liu and J. U. Kang, "Compressive SD-OCT: the application of compressed sensing in spectral domain optical coherence tomography," Opt. Express **18**(21), 22010–22019 (2010).
25. E. Lebed, P. J. Mackenzie, M. V. Sarunic, and M. F. Beg, "Rapid volumetric OCT image acquisition using compressive sampling," Opt. Express **18**(20), 21003–21012 (2010).
26. M. Young, E. Lebed, Y. Jian, P. J. Mackenzie, M. F. Beg, and M. V. Sarunic, "Real-time high-speed volumetric imaging using compressive sampling optical coherence tomography," Biomed. Opt. Express **2**(9), 2690–2697 (2011).
27. N. Mohan, I. Stojanovic, W. C. Karl, B. E. A. Saleh, and M. C. Teich, "Compressed sensing in optical coherence tomography," Proc. SPIE **7570**, 75700L, 75700L-5 (2010).
28. S. Jiao, R. Knighton, X. Huang, G. Gregori, and C. Puliafito, "Simultaneous acquisition of sectional and fundus ophthalmic images with spectral-domain optical coherence tomography," Opt. Express **13**(2), 444–452 (2005).
29. M. D. Robinson, S. J. Chiu, C. A. Toth, J. Izatt, J. Y. Lo, and S. Farsiu, "Novel applications of super-resolution in medical imaging," in *Super-Resolution Imaging,* P. Milanfar, ed. (CRC Press, 2010), pp. 383–412.
30. W. Dong, X. Li, L. Zhang, and G. Shi, "Sparsity-based image denoising via dictionary learning and structural clustering," in *2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (IEEE, 2011), pp. 457–464.
31. W. Dong, L. Zhang, and G. Shi, "Centralized sparse representation for image restoration," Proc. IEEE **ICCV**, 1259–1266 (2011).
32. S. Farsiu, M. Elad, and P. Milanfar, "A practical approach to superresolution," Proc. SPIE **6077**, 607703, 607703-15 (2006).
33. R. Estrada, C. Tomasi, M. T. Cabrera, D. K. Wallace, S. F. Freedman, and S. Farsiu, "Enhanced video indirect ophthalmoscopy (VIO) via robust mosaicing," Biomed. Opt. Express **2**(10), 2871–2887 (2011).
34. M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," IEEE Trans. Image Process. **15**(12), 3736–3745 (2006).
35. S. Li, L. Fang, and H. Yin, "An efficient dictionary learning algorithm and its application to 3-D medical image denoising," IEEE Trans. Biomed. Eng. **59**(2), 417–427 (2012).
36. J. Mairal, M. Elad, and G. Sapiro, "Sparse representation for color image restoration," IEEE Trans. Image Process. **17**(1), 53–69 (2008).
37. J. Mairal, G. Sapiro, and M. Elad, "Learning multiscale sparse representations for image and video restoration," Multiscale Model. Simulation **7**(1), 214–241 (2008).
38. A. M. Bruckstein, D. L. Donoho, and M. Elad, "From sparse solutions of systems of equations to sparse modeling of signals and images," SIAM Rev. **51**(1), 34–81 (2009).
39. M. Aharon, M. Elad, and A. M. Bruckstein, "The K-SVD: an algorithm for designing of overcomplete dictionaries for sparse representation," IEEE Trans. Signal Process. **54**(11), 4311–4322 (2006).
40. R. Rubinstein, M. Zibulevsky, and M. Elad, "Double sparsity: learning sparse dictionaries for sparse signal approximation," IEEE Trans. Signal Process. **58**(3), 1553–1564 (2010).
41. P. Chatterjee and P. Milanfar, "Clustering-based denoising with locally learned dictionaries," IEEE Trans. Image Process. **18**(7), 1438–1451 (2009).
42. W. Dong, L. Zhang, G. Shi, and X. Wu, "Image deblurring and super-resolution by adaptive sparse domain selection and adaptive regularization," IEEE Trans. Image Process. **20**(7), 1838–1857 (2011).
43. H. Takeda, S. Farsiu, and P. Milanfar, "Kernel regression for image processing and reconstruction," IEEE Trans. Image Process. **16**(2), 349–366 (2007).
44. B. Ophir, M. Lustig, and M. Elad, "Multi-scale dictionary learning using wavelets," IEEE J. Sel. Top. Signal Process. **5**(5), 1014–1024 (2011).

45. Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *1993 Conference Record of The Twenty-Seventh Asilomar Conference on Signals, Systems and Computers* (1993), Vol. 1, pp. 40–44.
46. A. Buades, B. Coll, and J.-M. Morel, "A review of image denoising algorithms, with a new one," Multiscale Model. Simul. **4**(2), 490–530 (2005).
47. 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).
48. S. Li and L. Fang, "Signal denoising with random refined orthogonal matching pursuit," IEEE Trans. Instrum. Meas. **61**(1), 26–34 (2012).
49. I. Daubechies, R. DeVore, M. Fornasier, and C. S. Gunturk, "Iteratively reweighted least squares minimization for sparse recovery," Commun. Pure Appl. Math. **63**(1), 1–38 (2010).
50. I. Daubechies, M. Defrise, and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," Commun. Pure Appl. Math. **57**(11), 1413–1457 (2004).
51. S. J. Chiu, J. A. Izatt, R. V. O'Connell, K. P. Winter, C. A. Toth, and S. Farsiu, "Validated automatic segmentation of AMD pathology including drusen and geographic atrophy in SD-OCT images," Invest. Ophthalmol. Vis. Sci. **53**(1), 53–61 (2012).
52. P. Thévenaz, U. E. Ruttimann, and M. Unser, "A pyramid approach to subpixel registration based on intensity," IEEE Trans. Image Process. **7**(1), 27–41 (1998).
53. W. Dong, "Sparsity-based Image Denoising via Dictionary Learning and Structural Clustering," *http://www4.comp.polyu.edu.hk/~cslzhang/code.htm*.
54. A. Foi, "Noise estimation and removal in MR imaging: The variance stabilization approach," in 2011 *IEEE International Symposium on Biomedical Imaging: from Nano to Macro* (2011), pp. 1809–1814.
55. G. Cincotti, G. Loi, and M. Pappalardo, "Frequency decomposition and compounding of ultrasound medical images with wavelet packets," IEEE Trans. Med. Imaging **20**(8), 764–771 (2001).
56. P. Bao and L. Zhang, "Noise reduction for magnetic resonance images via adaptive multiscale products thresholding," IEEE Trans. Med. Imaging **22**(9), 1089–1099 (2003).
57. M. D. Robinson, C. A. Toth, J. Y. Lo, and S. Farsiu, "Efficient fourier-wavelet super-resolution," IEEE Trans. Image Process. **19**(10), 2669–2681 (2010).
58. G. T. Chong, S. Farsiu, S. F. Freedman, N. Sarin, A. F. Koreishi, J. A. Izatt, and C. A. Toth, "Abnormal foveal morphology in ocular albinism imaged with spectral-domain optical coherence tomography," Arch. Ophthalmol. **127**(1), 37–44 (2009).
59. F. Luisier, T. Blu, and M. Unser, "A new SURE approach to image denoising: interscale orthonormal wavelet thresholding," IEEE Trans. Image Process. **16**(3), 593–606 (2007).
60. S. J. Chiu, X. T. Li, P. Nicholas, C. A. Toth, J. A. Izatt, and 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

In this paper, we introduce an image restoration method that attempts to recover the noiseless high-frequency information corrupted by the limitations of the state-of-the-art medical tomographic imaging systems. In particular, we focus on the spectral domain optical coherence tomography (SDOCT) systems [1,2], and show the applicability of our algorithm in reducing speckle noise [3,4].

Previous works in removing speckle noise can be categorized into two groups: model-based single-frame or multi-frame averaging techniques, as shown in Fig. 1. The first group utilizes classic denoising methods, which often assume an *a priori* parametric or non-parametric model for the signal and noise (e.g. Wiener filtering [5], kernel regression [6], or wavelets [7]). Theoretical bounds on the performance of such approaches are described in [8]. A few such algorithms already have been utilized for denoising ocular SDOCT images including [9–12] (Fig. 1(a)). The second group relies on capturing a sequence of repeated B-scans from a unique position. In a post-processing step, these images are registered and averaged, to create a less noisy image [13] (Fig. 1(b)). Some SDOCT imaging systems such as Spectralis (Heidelberg Engineering, Heidelberg, Germany) have a built-in image stabilization and averaging system and can directly produce high-SNR images. For others, such as Bioptigen SDOCT system (Bioptigen Inc., Research Triangle Park, NC), registration and averaging is done post image capturing phase. Indeed, for either of these SDOCT systems, averaging based denoising dramatically increases the image acquisition time. To reduce image acquisition time, a very recent work [14] utilizes wavelets for denoising a sequence of repeated SDOCT B-scans captured from a unique position.
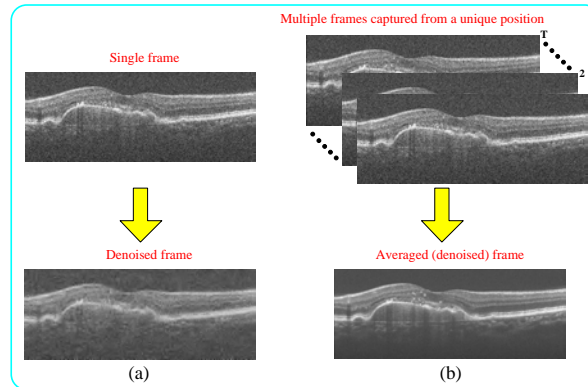
Fig. 1. Two classic methods for removing noise in SDOCT image. (a) Denoising by a model-based single-frame technique [7]. (b) Denoising by multi-frame averaging technique: SDOCT frames acquired from the same location are averaged to increase SNR.

In this work, we present a novel hybrid approach which is called multiscale sparsity based tomographic denoising (MSBTD) for denoising volumetric SDOCT scans, where the imaged volume is sampled at several azimuthally distanced B-scans (Fig. 2(a)). The MSBTD method utilizes a non-uniform scanning pattern, in which, a fraction of B-scans are captured slowly at a relatively higher than nominal signal-to-noise ratio (SNR). The rest of the B-scans are captured fast at the nominal SNR. Utilizing compressive sensing principles [15–21], we learn a sparse representation dictionary for each of these high-SNR images and utilize these dictionaries to denoise the neighboring low-SNR B-scans. The rationale for this approach is that neighboring B-scans, in common SDOCT volumes, are expected to have similar texture and noise pattern, as illustrated in Fig. 2. The exciting property of our approach is that, unlike [14], it does not require capturing more than one B-scan from the majority of azimuthal positions and therefore, it requires significantly less scanning time. Moreover, the MSBTD method is well suited to retrospectively improve the quality of images in databanks from large scale multi-center clinical trials [22]. Note that, traditionally, ocular SDOCT imaging protocols often require at least two types of SDOCT scans: the first, a densely sampled (but low-SNR) large field-of-view volumetric scan and the second, sparsely sampled high-SNR scans (often consisting of only one image from the fovea). In section 3, we show the applicability of the MSBTD method for such data sets.
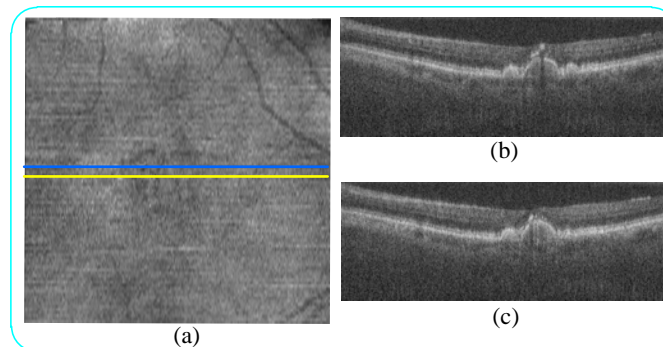

Fig. 2. In common scanning patterns, SDOCT B-scans acquired from adjacent (azimuthally distanced) positions (a) have similar features (b,c). (a) In the summed-voxel projection [28,29] (SVP) *en face* SDOCT image of a non-neovascular age-related macular degeneration (AMD) patient, (b,c) B-Scans are acquired from the location of the blue and yellow lines, respectively.

A few previous works have studied related algorithms in the context of photonics imaging. The first hardware implemented experiment demonstrating the superiority of adaptive sampling in improving the scanning speed and SNR of ballistic photon based imagers was reported in [23]. More recently, application of compressive sensing theory for speeding up the SDOCT scanning time has attracted sizable attention [24–27]. While these important papers have described different implementations of compressive sampling in SDOCT imaging, they provide limited performance analysis, as their experimental results have been only compared to basic none-data-adaptive techniques such as cubic interpolation. A fundamental question, i.e. "Is sparsity based reconstruction of SDOCT images beneficial as compared to alternative modern adaptive restoration/reconstruction techniques?" has yet to be answered.

To address this question, in Section 2, we introduce the MSBTD algorithm based on irregular sampling and sparse restoration, and demonstrate its superiority compared to popular modern adaptive denoising algorithms. While the sparse representation objective function for the MSBTD method [30,31] (as with virtually all other works in this field) has been known to the image processing community, the main novelty of our algorithm is its integration with a customized scanning algorithm, which adapts it for clinical SDOCT applications. As noted, an interesting aspect of our paper is its universal application. To prove this point, in Section 3, unlike many other previous works in this field, we abstain from experimenting on simulated or controlled data sets captured in the laboratory environment, which might be biased or unfairly optimized in favor of the proposed algorithm. Instead, we test the MSBTD algorithm using an unbiased randomly selected deidentified data set from a multi-center clinical trial, the imaging protocol of which was designed years ago, independent of the algorithm in this paper. Concluding remarks are given in Section 4, discussing future possible applications of the proposed technique.

2. Methods

In our previous works, we have demonstrated the application of multi-frame image fusion methods to overcome the theoretical and practical constraints (e.g. resolution [32] and field-of-view [33]) of modern optical imaging systems. As briefly reviewed in the previous section, the proposed MSBTD algorithm is based on a customized scanning pattern, in which, a minority number of B-scans are captured at higher SNR compared to others. We use these high-SNR images to learn a sparse model of underlying anatomical structure in this sequence of B-Scans and apply this model to denoise the low-SNR images.

2.1. Sparse representation of SDOCT images

A volumetric SDOCT scan, $\mathbf{Y} \in \mathbb{R}^{N \times M \times L}$, consists of L images (B-scans), each representing a 2D slice, $\mathbf{Y}_l \in \mathbb{R}^{N \times M}$, of a 3-D structure. In many clinical applications, captured B-scans are very noisy and require image enhancement (denoising) before any visual or quantitative analysis.

In recent years, many compressive sensing/sparse representation based algorithms have been proposed as promising alternative approaches to classic denoising techniques [31,34–37]. Sparse representation [38] models noiseless signal as a sparse linear combination of basis elements, termed atoms, from a dictionary.

We represent a small rectangular patch (of size $N' \times M'$) in the desired noiseless image as $\mathbf{X}$. A vector representation of this patch can be obtained by lexicographic ordering: $\mathbf{x} \in \mathbb{R}^Q; Q = N' \times M'$. This vector can be synthesized by a linear combination of basis functions assembled in a matrix called dictionary, $\mathbf{D} \in \mathbb{R}^{Q \times P}$, as

$$\mathbf{x} = \mathbf{D}\alpha, \tag{1}$$

where $P>Q$ implies that the dictionary is redundant and $\boldsymbol{\alpha} \in \mathbb{R}^P$ is a vector of scalar coefficients. The basic idea of compressive sensing is that when for a class of images the majority of elements in vector $\boldsymbol{\alpha}$ corresponding to a particular dictionary are zeros or very small (sparsity), relatively few well-chosen measurements suffice to reconstruct the images in this class [21].

We denote the patch contaminated with noise as $\mathbf{x}^{Noise} \in \mathbb{R}^Q$. Since, unlike the SDOCT image, noise is not expected to be sparsely represented by $\mathbf{D}$, the sparse solution of the denoising problem [34] can be obtained from

$$\hat{\boldsymbol{\alpha}} = \arg \min_{\alpha} \left\| \boldsymbol{\alpha} \right\|_0 \ \text{ subject to } \ \left\| \mathbf{x}^{Noise} - \mathbf{D}\boldsymbol{\alpha} \right\|_2 \le \omega, \tag{2}$$

where $\left\| \bullet \right\|_2^2$ and $\left\| \bullet \right\|_0$ represent the ℓ_2 and ℓ_0 -norms, respectively, and ω is the error tolerance.

A fundamental consideration in employing the above model is the selection of the dictionary $\mathbf{D}$. One popular class of sparsity based denoising algorithms exploits the information of the noisy image itself to define the dictionary (which we denote as $\mathbf{D}^{Noise}$) [34]. While for many situations this method provides very promising results, however, the high-level of noise in the SDOCT images negatively interferes with the learning process, degrades the quality of the trained dictionary, and subsequently results in a suboptimal image restoration. An alternative (ideal) approach is to learn the dictionary from the noiseless image. Since in practice, such an ideal image is not available, we rely on a less noisy image $\mathbf{Y}_l^{Ave}$, obtained from registering and averaging a sequence of (e.g. T) repeated B-scans $\mathbf{Y}_l^1,...,\mathbf{Y}_l^t,...,\mathbf{Y}_l^T$ from a unique position (Fig. 1). It is natural to assume that the dictionary trained on this averaged image, $\mathbf{D}^{Ave}$, is less affected by noise as compared to $\mathbf{D}^{Noise}$.
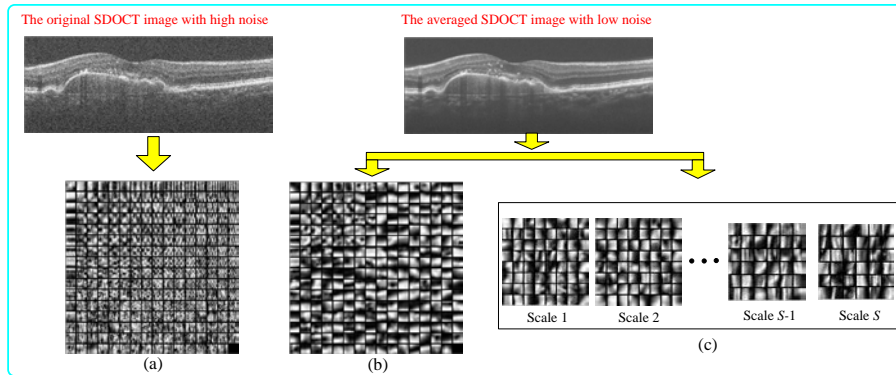

Fig. 3. Dictionaries trained on (a) the original low-SNR SDOCT image by the K-SVD algorithm, (b) averaged high-SNR image by the K-SVD algorithm, and (c) averaged high-SNR image by the proposed MSBTD training algorithm (due to the limited space, we show only the first atom of each learned subdictionary).

To directly compare these two methods, we use the popular K-SVD training algorithm [39] and the proposed algorithm described in the following subsection for the dictionary learning process. Figure 3(a) shows an illustrative example of a dictionary trained using the K-SVD algorithm on a low-SNR B-scan. Figure 3(b,c) show corresponding examples of dictionaries trained using the K-SVD and the MSBTD algorithms on an averaged high-SNR image, respectively. Visual inspection of the dictionary $\mathbf{D}^{Noise}$ suggests that it is more affected by noise as compared to $\mathbf{D}^{Ave}$. Ergo, in contrast to [34], we propose to denoise each low-SNR image utilizing a dictionary learned from a nearby (or even comparatively distant) averaged (high-SNR) image. This strategy allows us to reduce the noise disturbance in the dictionary

learning process, which we expect to enhance the denoising outcome, without significantly increasing the image acquisition time. In the experiment section, we quantitatively compare the performance of such dictionaries in denoising SDOCT images.

2.2. Multiscale structural dictionary

As noted, in compressive sensing, design of the dictionary is a critical and often complex issue. Applying K-SVD on the whole image is computationally very expensive. An alternative for reducing computational complexity is to cut the training image into overlapping fixed-size patches, and learn a universal dictionary $\mathbf{D}$ on these patches by the K-SVD algorithm [34] or its variants [35,40]. Since such a universal dictionary is neither optimal nor efficient to represent different structures, we alternatively learn a set of subdictionaries $\left\{ \mathbf{D}_k^{Str} \in \mathbb{R}^{Q \times Q} \right\}$, $k = 1, 2, ..., K$, each best fit a particular structure [41,42]. This is achieved by first clustering the training patches into K structural clusters using the k-means approach. We will use the centroid of each cluster ($\mathbf{c}_k \in \mathbb{R}^Q$) in a later dictionary selection step (Section 2.3.1). Next, a subdictionary is deduced from each of the K clusters using the principle component analysis (PCA) algorithm.

To effectively capture the properties of different structures and textures on ocular SDOCT images (e.g. noting that each retinal layer has a different thickness) with the learned dictionary, it is natural to prefer variable sized training patches. This is analogous to the adaptive kernel shape and size discussion in the classic image denoising literature (smaller kernels are preferred in the texture area, while larger kernels are used in the smooth area) [43]. Both approaches [41,42] noted in the above paragraph use fixed size patches, since neither of k-means and PCA algorithms (in their primal forms) can deal with the data of varying size. To address this problem, the idea of learning multiscale dictionary has been introduced in two recent works [37,44]. These methods adopt the K-SVD algorithm to learn subdictionaries on each image domain obtained from the wavelet or quadtree decomposition. However, the wavelet or quadtree decomposition can only be regarded as a way to downsample the training image (zooming out). One may argue that such approaches might not efficiently represent detailed features, which are smaller than the base (zeroth-scale) patch (zooming in).
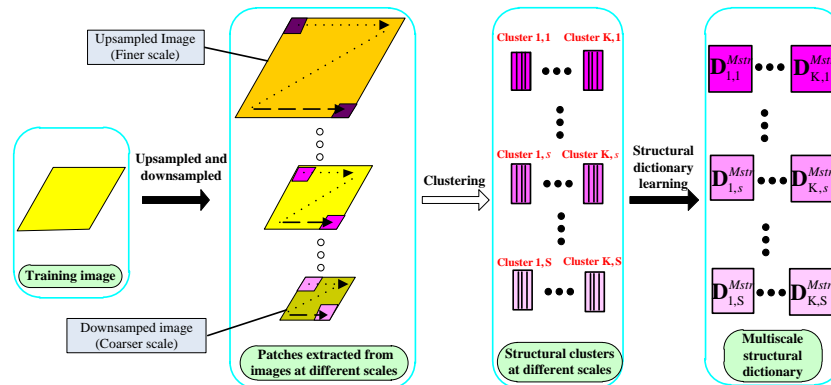

Fig. 4. Algorithmic flowchart of multiscale structural dictionary learning process.

To resolve this problem, we incorporate a basic multiscale approach into the structural learning process. Specifically, instead of varying the patch size, the training image itself is first zoomed in and out through upsampling and downsampling processes. Then, the original and these magnified images are divided into same sized patches. Ergo, although the patch size is fixed, patches from different magnified scales essentially can be considered as variable sized patches from a particular scale. Next, the structural learning process is applied on the

patches from the same scale (s) to create the multiscale structural dictionary, which is the concatenation of the subdictionaries ($\left\{\mathbf{D}_{k,s}^{Mstr}\right\}, s = 1, 2, ..., S$) from all scales. Figure 4 shows a schematic representation of the multiscale structural dictionary learning process. In this paper, we implemented the upsampling and downsampling processes by the bilinear interpolation. Indeed, to improve performance, one may use more advanced image resizing algorithms, which may result in improved performance, while increasing computational complexity.

2.3. Nonlocal denoising process

The previous subsection described a method to learn a multiscale structural dictionary from a high-SNR image. This subsection describes how to utilize this learned dictionary for denoising low-SNR SDOCT images. As in the dictionary learning step, we divide the low-SNR image $\mathbf{Y}_l$ into overlapping patches $\mathbf{y}_1, ..., \mathbf{y}_i, ..., \mathbf{y}_I$ of size $w \times z$, where the distance between each patch is V pixels in each direction (i.e. for an image of size $N \times M$, the number of the extracted patches (I) is $(N - w) \times (M - z)/(V \times V)$). For each patch ($\mathbf{y}_i$), we find an appropriate subdictionary $\mathbf{D}_i^A$ from the learned multiscale structural dictionary to sparsely represent the patch, which implicitly results in denoising of that patch. These steps are detailed in the following.

2.3.1. Subdictionary selection

To find the best subdictionary among the collection of learned subdictionaries for each noisy patch, we compare representative features of the patch ($\mathbf{y}_i$) and each subdictionary. To represent each subdictionary, we use the centroid atom ($\mathbf{c}_{k,s} \in \mathbb{R}^{(w \bullet z) \times 1}$) of the corresponding k-means cluster (Section 2.2) [42]. As for the representative feature of each patch, we utilize its high-frequency component (denoted by $\mathbf{y}_i^{Hf}$) [42]. We find the best fitted subdictionary $\mathbf{D}_i^A$ for the patch $\mathbf{y}_i$ based on the normalized correlation [35,45] between $\mathbf{c}_{k,s}$ and $\mathbf{y}_i^{Hf}$:

$$A = (k_i, s_i) = \arg \max_{k,s} \left| \left\langle \mathbf{c}_{k,s}, \mathbf{y}_i^{Hf} \right\rangle \right|. \tag{3}$$

2.3.2. Sparse representation of patches

After finding the appropriate subdictionary, following [31], we define the objective function for the sparse representation of the noisy patch as follows:

$$(\hat{\boldsymbol{\alpha}}_i, \hat{\boldsymbol{\beta}}_i) = \arg \min_{\boldsymbol{\alpha}_i, \boldsymbol{\beta}_i} \left\{ \left\| \mathbf{y}_i - \mathbf{D}_i^A \boldsymbol{\alpha}_i \right\|_2^2 + \lambda_1 \left\| \boldsymbol{\alpha}_i \right\|_0 + \lambda_2 \left\| \boldsymbol{\alpha}_i - \boldsymbol{\beta}_i \right\|_0 \right\}, \tag{4}$$

where λ_1 and λ_2 are scalar Lagrange multipliers and $\boldsymbol{\alpha}_i$ is a sparse vector of coefficients for the ith denoised patch. The sparse vector $\boldsymbol{\beta}_i$ is the expected value of a sparse vector representing the corresponding noiseless patch. While the first regularization term allows us to exploit local sparsity in a patch, the second regularization term (through $\boldsymbol{\beta}_i$) allows us to utilize the non-local similarities (data redundancy) often seen in natural images [31,35,46,47]. In practice, the noiseless patch is approximated as the average of J patches with the highest similarity to the processed patch $\mathbf{y}_i$ within a searching window [31].

The double-head optimization problem (4) is solved by the iterative reweighted algorithm in [31]. In each iteration, one of the unknown parameters is kept constant and the other can be efficiently found by greedy pursuit [45,48] or iterative shrinkage algorithms [49,50], which replace the ℓ_0-norm with the ℓ_1-norm. After finding the sparse representation of all (I)

patches ($\hat{\boldsymbol{\alpha}}_1,...,\hat{\boldsymbol{\alpha}}_i,....,\hat{\boldsymbol{\alpha}}_I$), the estimate of the denoised patches can be easily obtained from $\mathbf{D}_1^A\hat{\boldsymbol{\alpha}}_1,...,\mathbf{D}_i^A\hat{\boldsymbol{\alpha}}_i,....,\mathbf{D}_I^A\hat{\boldsymbol{\alpha}}_I$. Then, we reformat the estimated patches from lexicographic vectors to rectangular shape and return them to their original geometric position. Since these patches are highly overlapped, we average their overlapping sections to reconstruct the denoised image as described in [34]. The outline of the whole denoising process is shown in Fig. 5.
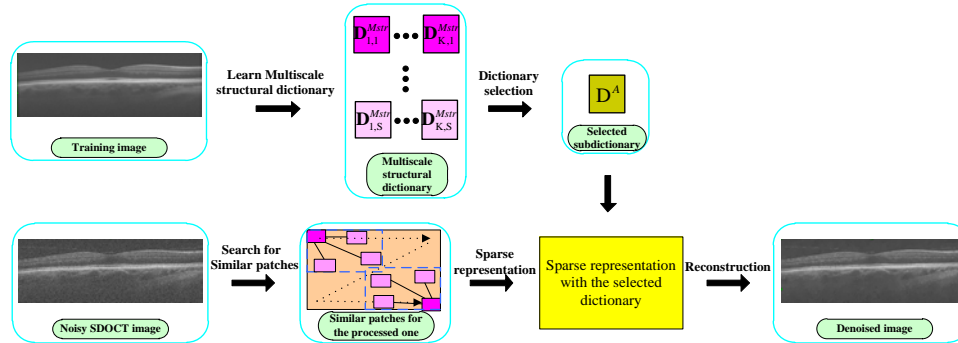

Fig. 5. The outline of the MSBTD algorithm.

2.4. Algorithmic parameters

For the specific case of denoising retinal SDOCT images, we have empirically selected algorithmic parameters, which are kept unchanged for all experiments in this paper. Since most of the similar structures in the SDOCT image lie on the horizontal direction, the patch and the search window sizes are chosen to be rectangle of size 3×20 and 40×60 pixels, respectively. The distance V between each processed patch is 1 and the number J of similar patches in each searching window is 20. In the k-means clustering stage, the cluster number K is set to 70. Prior to the multiscale learning process, the original image is upsampled two times, each by a factor 1.25 and downsampled three times, each by a factor of 1.5625 to create the training images of six scales. The parameters of the iterative reweighted algorithm for solving the problem (4) are set to the default values in [31].

2.5. Data sets

We tested our algorithm on randomly selected data sets from 17 eyes from 17 subjects with and without nonneovascular age-related macular degeneration (AMD) enrolled in the A2A SDOCT study, which was registered at clinicaltrials.gov [22] and approved by the institutional review boards (IRBs) of the four A2A SDOCT clinics (Devers Eye Institute, Duke Eye Center, Emory Eye Center, and the National Eye Institute). In particular, ten data sets are from normal subject, while the rest are from AMD subjects. The study complied with the Declaration of Helsinki, and informed consent was obtained from all participants.

All data sets were captured before the start of this project and the imaging protocol was not altered in any form for this study. In the A2A SDOCT study, volumetric scans were acquired using SDOCT imaging systems from Bioptigen, Inc. (Research Triangle Park, NC). For each patient across all sites, two types of SDOCT scans were captured. 1) A square (~6.6 × 6.6mm) volume scan with 1000 A-scans and 100 B-scans, which included the fovea. The scan sizes and the axial, lateral, and azimuthal resolutions varied slightly by site, and are specified in Table 1 of [51]. 2) An azimuthally repeated scan with 1000 A-Scans and 40 B-Scans targeted at the fovea. We cropped each A-scan, to achieve B-Scans of size 280×1000, to only include the informative areas (excluding the smooth dark areas deep below the choroid or in the vitreous).

2.6. Data processing

We registered the azimuthally repeated scan using the ImageJ (software; National Institutes of Health, Bethesda, Maryland, USA) StackReg Registration plug-in [52]. We implemented all other image processing tasks in MATLAB (MathWorks, Natick, MA). To implement the iterative reweighted algorithm, we used the code provided in [53]. In addition, we employed the MATLAB code in [54] to estimate the noise variance in the test images.

2.7. Quantitative measures of performance

We used mean-to-standard-deviation ratio (MSR) [55], contrast-to-noise ratio (CNR) [56], and peak signal-to-noise-ratio (PSNR) to compare the performance of different denoising algorithms. The metrics MSR and CNR are defined as follows:

$$\text{MSR} = \frac{\mu_f}{\sigma_f}, \tag{5}$$

$$\text{CNR} = \frac{|\mu_f - \mu_b|}{\sqrt{0.5(\sigma_f^2 + \sigma_b^2)}}, \tag{6}$$

where μ_b and σ_b are the mean and the standard deviation of the background region (e.g. red box #1 in Fig. 6), while μ_f and σ_f are the mean and the standard deviation of the foreground regions (e.g. red box #2-6 in Fig. 6).

As a global model of image quality, we used the peak signal-to-noise-ratio (PSNR) [57], defined as

$$\text{PSNR} = 20 \cdot \log_{10} \left(\frac{\text{Max}_{\mathbf{R}}}{\sqrt{\frac{1}{H} \sum_{h=1}^{H} \left(\mathbf{R}_h - \hat{\mathbf{R}}_h \right)^2}} \right), \tag{7}$$

where $\mathbf{R}_h$ is the hth pixel in the reference noiseless image $\mathbf{R}$, $\hat{\mathbf{R}}_h$ represents the hth pixel of the denoised image $\hat{\mathbf{R}}$, H is the total number of pixels, and $\text{Max}_{\mathbf{R}}$ is the maximum intensity value of $\mathbf{R}$.

Since there is no ideal "noiseless" image available for these clinical experiments, in Experiment 1, we used the averaged foveal image as a noiseless approximation to the corresponding foveal B-scan from the noisy volumetric scan. We identified the low-SNR foveal B-scan corresponding to the averaged scan by visual comparison. To have the best matching, the averaged and noisy images were registered using the StackReg Registration plug-in for ImageJ [52]. We note that in practice, such visual comparison is not necessary and the performance of our algorithm is independent of information about the location of the averaged high-SNR B-scan. We use the averaged B-scan location only to generate quantitative performance metrics, i.e. PSNR. In Experiment 2, we tested on more distant B-scans for which averaged (noiseless) image were not available. In these cases, we only reported MSR and CNR values.

3. Experimental results

To evaluate the effectiveness of the proposed MSBTD method, we compared its performance with those from four well-known denoising approaches: Tikhonov [58], New SURE [59], K-SVD [34], and BM3D [47]. In these experiments, the parameters for the Tikhonov [58] method are tuned to reach the best results, while the parameters of the New SURE, K-SVD,
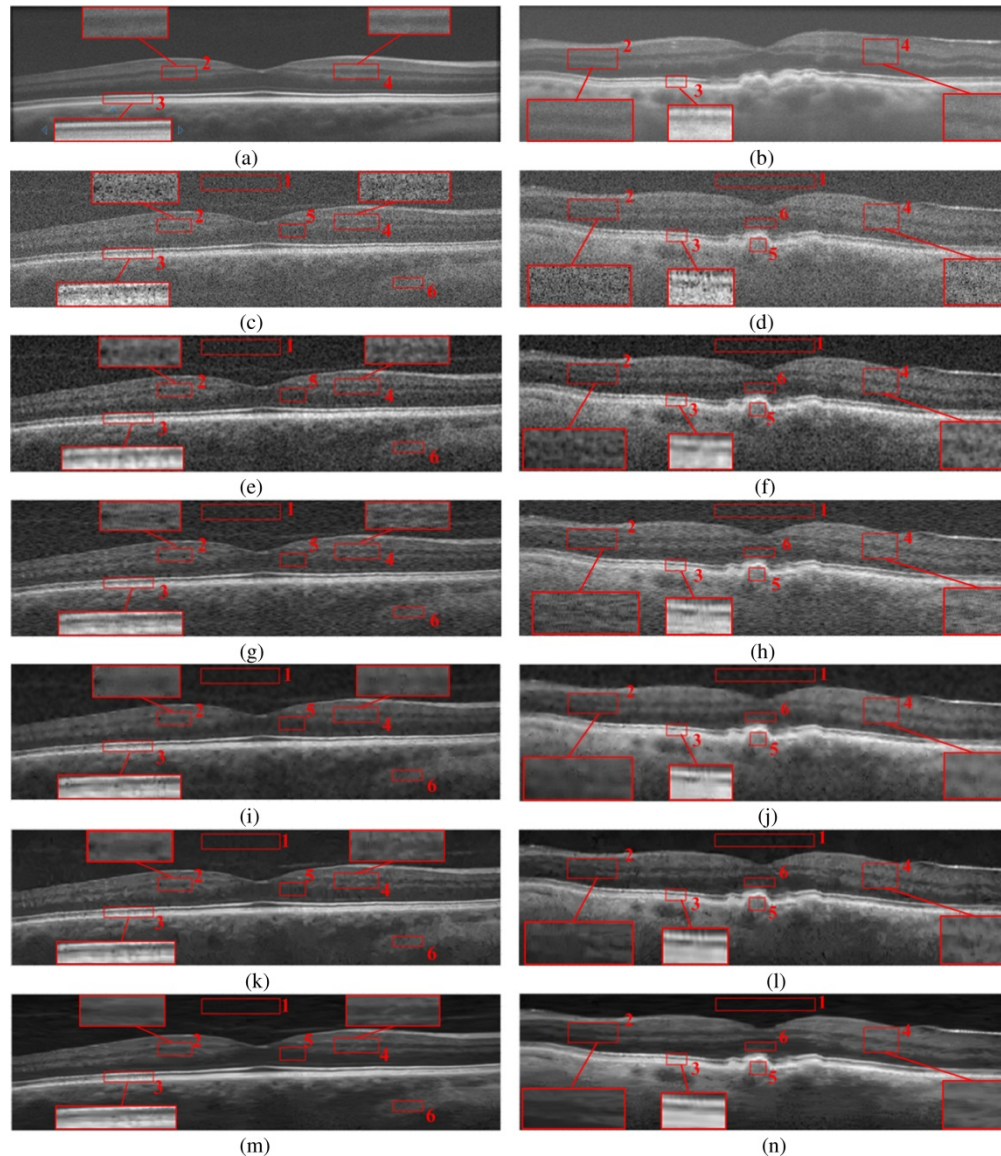
Fig. 6. Denoising results for two test SDOCT retinal images using the Tikhonov [58], New SURE [59], K-SVD [34], BM3D [47], and the proposed MSBTD method. The left and right columns show the foveal images from a normal subject and an AMD patient, respectively. (a, b) The averaged (high-SNR) images for the multiscale structural dictionary training. (c, d) Low-SNR noisy foveal images from the volumetric scans. (e, f) Denoising results using the Tikhonov method. (g, h) Denoising results using the New SURE method. (i, j) Denoising results using the K-SVD method. (k, l) Denoising results using the BM3D method. (m, n) Denoising results using the MSBTD method.

and BM3D methods are set to the default values as in their corresponding references [34,47,59].

3.1. Experiment 1: denoising based on learned dictionary from a nearby high-SNR Scan

Figure 6. shows qualitative comparisons of two raw SDOCT retinal images (from a normal and an AMD subject) and their denoised versions obtained from the noted denoising methods. The raw (low-SNR) images are the geometrically closest (less than 66µm distanced) B-scans

to the averaged (high-SNR) images. Since the boundaries between retinal layers contain important anatomic and pathologic information [60], three boundary regions (boxes #2, 3, 4) in these images are marked with red rectangle and magnified. As can be observed, the Tikhonov [58] and New SURE [59] methods provide limited noise suppression for the test images. Although the K-SVD method better suppresses the noise, it introduces over-smoothing, thus leading to significant loss of image details. The BM3D method delivers improved noise suppression and limits the over-smoothing problem to some extent, but creates obvious artifacts. By contrast, application of the proposed MSBTD method results in noticeably improved noise suppression, while preserving details as compared to other methods. Especially, note in the layer boundary preservation in the area magnified by the red boxes (e.g. #2 and #4).

To compare these methods in a quantitative manner, we measured MSR and CNR on six regions of interest (similar to the red boxes #1-6 in Fig. 6) from 17 test images of different (randomly selected) subjects. In each image, we averaged the MSR and CNR values obtained for the five foreground regions (e.g. red box #2-6 in Fig. 6). We report the mean and standard deviation of these averaged MSR and CNR results across all the test images in Table 1.

Table 1. Mean and standard deviation of the MSR and CNR results for seventeen SDOCT retinal images using the Tikhonov [58], New SURE [59], K-SVD [34], BM3D [47], and the proposed MSBTD method. The best results in the table are shown in bold.

	Original	Tikhonov [58]	New SURE [59]	K-SVD [34]	BM3D [47]	MSBTD
Mean (CNR)	1.27	3.13	2.49	4.11	3.89	**4.76**
Standard deviation (CNR)	0.43	0.94	0.60	1.23	1.05	1.54
Mean (MSR)	3.20	7.62	6.74	11.22	11.52	**14.76**
Standard deviation (MSR)	0.46	0.95	1.69	2.77	2.42	4.75

Next, we compared the PSNR of these methods for all the seventeen subjects. The mean and standard deviation of the PSNR results are reported in Table 2. As for the case of MSR and CNR shown in Table 1, we observe that the MSBTD method delivers the best results in the PSNR sense. Furthermore, we note that although the PSNR of the K-SVD is close to that of the MSBTD method, our clinical collaborators preferred the visual outcome of the MSBTD method, as illustrated in Fig. 7.

Table 2. Mean and standard deviation of the PSNR (dB) for seventeen SDOCT foveal images obtained from the Tikhonov [58], New SURE [59], K-SVD [34], BM3D [47], and the proposed MSBTD method. The best result in the table is shown in bold.

	Tikhonov [58]	New SURE [59]	K-SVD [34]	BM3D [47]	MSBTD
Mean (PSNR)	23.67	23.46	26.13	26.04	**26.46**
Standard deviation (PSNR)	0.96	1.40	1.70	1.65	1.72

3.2. Experiment 2: denoising based on learned dictionary from a distant high-SNR scan

In the previous experiments, to denoise the test images, we trained the multiscale structural dictionary on a geometrically close (adjacent) averaged image. To test whether the learned dictionary is effective for denoising any other macular scan (i.e. distanced up to 3.3mm from the fovea), we apply a single learned dictionary to images captured from four different locations and compare the results of the MSBTD with those of the Tikhonov [58], New SURE

[59], K-SVD [34], and BM3D [47]. The quantitative results are reported in Table 3. Similar to Experiment 1, six regions (similar to the red boxes #1-6 in Fig. 8) in each image are selected to compute the CNR and MSR metrics (for this case, we could not report PSNR since no valid denoised image is available). We observe that compared to the other methods, the MSBTD method still achieves the best performance in terms of the quantitative measures (CNR and MSR). Furthermore, we provide the qualitative comparison of the MSBTD method with the K-SVD and BM3D in Fig. 8. For easy comparison, three boundary regions (boxes #2, 3, 4) in each image are magnified as well. As can be observed, the visual appearance of the MSBTD is superior to that of the K-SVD and BM3D (and significantly better than the weaker methods Tikhonov [58], New SURE [59], results of which omitted to save space).

Table 3. Mean and standard deviation of the MSR and CNR measures for four test images acquired from different positions using the Tikhonov [58], New SURE [59], K-SVD [34], BM3D [47], and the MSBTD method. The best results in the table are shown in bold.

	Original	Tikhonov [58]	New SURE [59]	K-SVD [34]	BM3D [47]	MSBTD
Mean (CNR)	1.18	3.26	2.81	5.00	4.93	**6.04**
Standard deviation (CNR)	0.17	0.22	0.25	0.04	0.39	0.71
Mean (MSR)	3.22	7.64	7.58	12.73	12.96	**16.32**
Standard deviation (MSR)	0.08	0.63	0.48	1.86	2.52	3.95

On average, denoising one B-scan required about 9 minutes, implemented in MATLAB code executed on a desktop computer with an Intel (R) Core i7-2600K CPU and 8 GB of RAM. Our code was not optimized for speed. The processing time is expected to be reduced significantly by more efficient coding coupled with a general purpose Graphics Processing Unit (GPU) [26].

4. Discussion and Conclusions

In this paper, we demonstrated a novel approach called the MSBTD for denoising SDOCT images using the sparse representation technique. We proposed to capture SDOCT volumes with a non-uniform scanning pattern, where a selected number of B-scans were imaged at higher SNR. We learned a sparse representation dictionary for each of these high-SNR images, and utilized them to denoise the low-SNR B-scans. We demonstrated that the MSBTD method outperforms state-of-the-art denoising methods. To encourage further research in this area, we have made the data set and the software that we have developed for this project publically available at http://www.duke.edu/~sf59/Fang_BOE_2012.htm.

In the experimental section, we selected the parameters for the patch and searching window empirically, and kept them fixed. Note that, the optimal choices for such parameters should depend on the specific application. For example, as the search window size increases, the denoising performance of the MSBTD method is expected to slightly improve. However, this improvement in performance comes with the increase in the computational complexity. On the other hand, utilizing larger patches suppresses noise more efficiently but results in a more aggressive smoothing and the loss of image details.

An interesting aspect of the MSBTD algorithm is that we only needed to capture very few (in our experiments only one) high SNR image for denoising the full volume. We believe the reason for this phenomenon is that while retinal SDOCT images from different locations in the eye may appear very different, they are all built with similar building blocks (multiscale structural dictionary representing retinal layers with different thicknesses and orientations). Moreover, we were lucky that our high-SNR images are commonly captured from the fovea, in which, retinal layers have a different thickness and orientating in the central region as
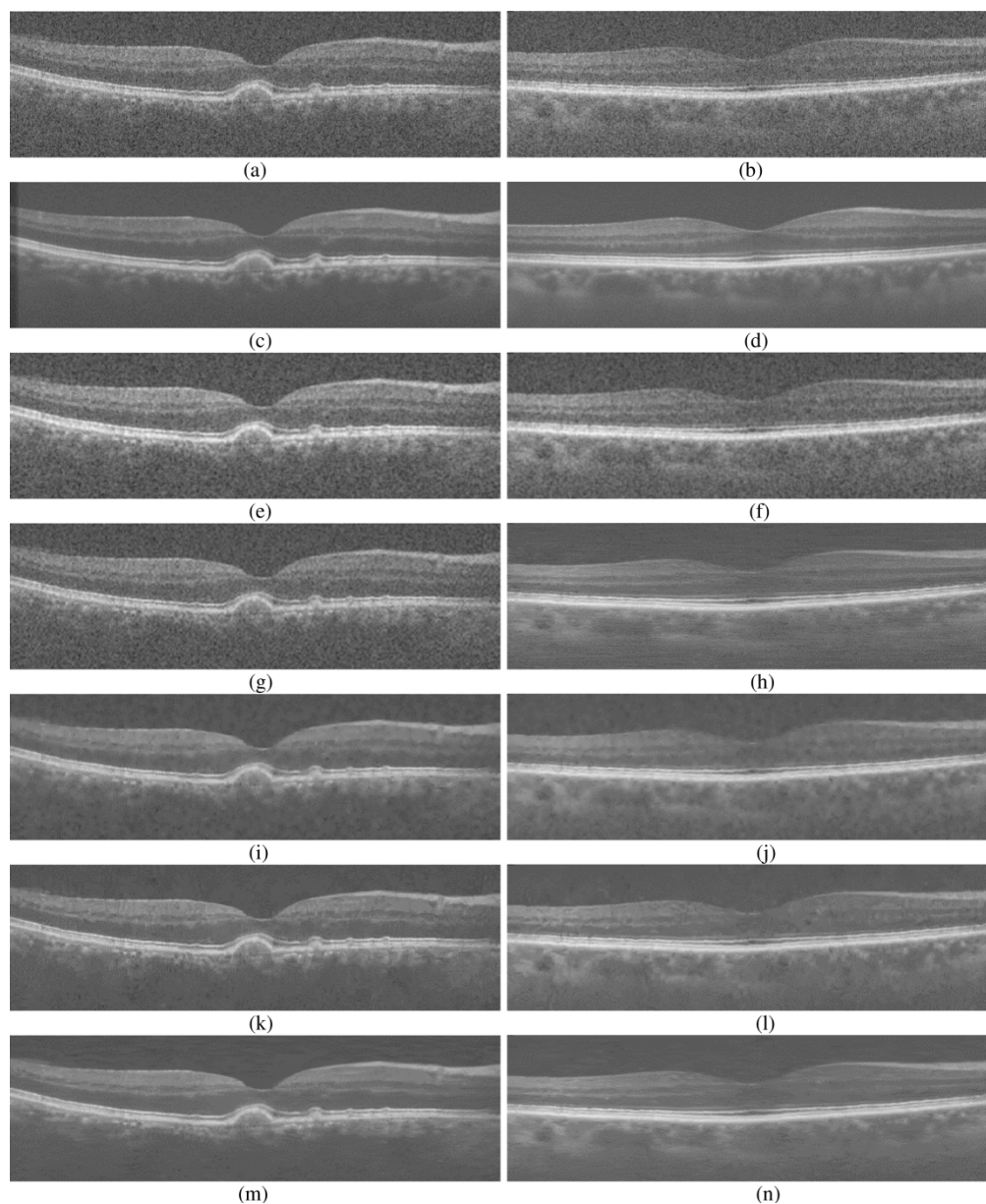
Fig. 7. Visual comparison of denoising results from the Tikhonov [58], New SURE [59], K-SVD [34], BM3D [47], and MSBTD on two SDOCT test images. (a, b) Test SDOCT images. (c, d) Averaged images. (e, f) Denoising results using the Tikhonov method (Left: PSNR = 22.67, Right: PSNR = 24.01). (g, h) Denoising results using the New SURE method (Left: PSNR = 25.39, Right: PSNR = 25.35). (i, j) Denoising results using the K-SVD method (Left: PSNR = 27.98, Right: PSNR = 25.81). (k, l) Denoising results using the BM3D method (Left: PSNR = 27.72, Right: PSNR = 25.69). (m, n) Denoising results using the MSBTD method (Left: PSNR = 28.19, Right: PSNR = 26.06).
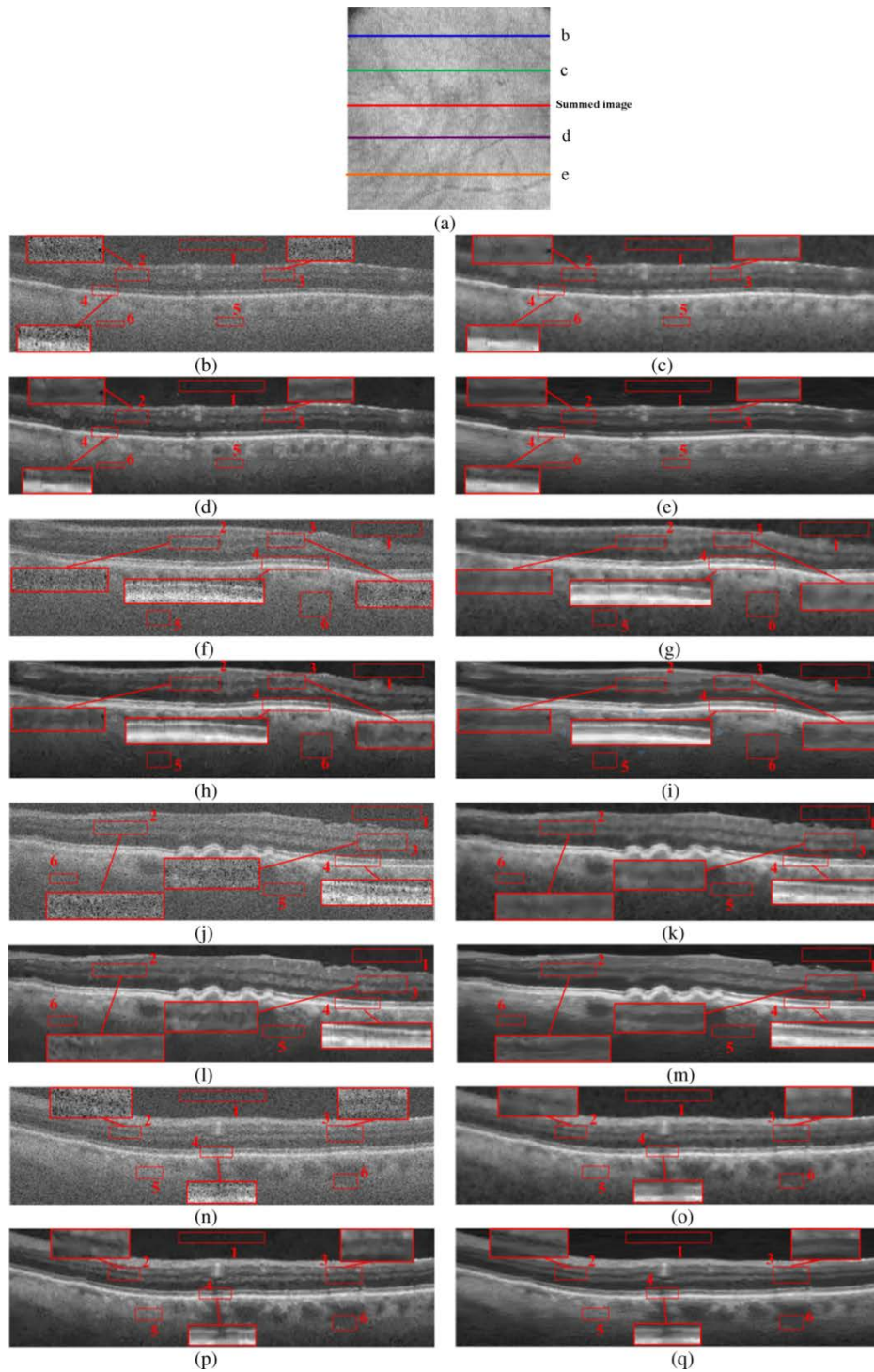
Fig. 8. Denoising results for test images captured from four different locations processed by the same learned dictionary. (a) Summed-voxel projection (SVP) *en face* image of an AMD eye (~6.6mm^2). (b, f, j, n) Raw noisy images acquired from the location b-e shown in (a). (c, g, k, o) Denoising results using the K-SVD [34]. (d, h, l, p) Denoising results using the BM3D [47]. (e, i, m, q) Denoising results using the MSBTD.

compared to the periphery. Such diversity in the structure of foveal B-scan results in more efficient and generalizable learned sub-dictionaries. This is important for two reasons. First, it means that the total scanning time for capturing a high-SNR volume can be reduced by utilizing the MSBTD method. Second, we can retrospectively denoise SDOCT scans from many other clinical trials, which use at least two very common scanning protocols, i.e. the dense volumetric low-SNR scan, and the so-called limited linear (averaged) high-SNR scan.

We note that the foveal (averaged) image sequence was captured in less than two seconds. This short scanning time reduces the probability of large motion artifacts. In the rare case that some of these frames are corrupted by artifacts like blinking, such frames can be easily detected by a simple normalized cross-correlation test [57] and removed from the averaging set.

Another interesting case is when the volumetric scan pattern is very dense in the azimuthal direction and the neighboring B-scans might have similar structures. Such an approximation is more valid in the normal eyes (as compared to the AMD eyes). In such cases, if there is no access to the averaged frames, we can approximate the training high-SNR image by averaging a few neighboring B-scans, which are away from the fast changing foveal region.

In this paper, we demonstrated the applicability of our algorithm for denoising images from the normal and nonneovascular AMD subjects. In our future publications, we will address the applicability of our denoising method for efficient image analysis in other types of retinal diseases including the diabetic macular edema. In addition, it is worthy to note that while we used this algorithm for denoising purposes, it is straightforward to use it for many other image restoration/enhancement applications such as deblurring, interpolation, and super-resolution. Moreover, the proposed MSBTD method is not necessarily the optimal compressive sensing approach for denoising SDOCT images. Our method was based on many on-the-shelf algorithms originally designed for other models of signal and noise. A more efficient adaptation of the compressive sensing technique to the ocular SDOCT signal and noise properties and its utilization for a variety of inverse problem applications are parts of our ongoing work.

Acknowledgments