3-D OCT Data Denoising with Nonseparable Oversampled Lapped Transform

Shogo MURAMATSU*, Samuel CHOI* and Takumi KAWAMURA†
* Dept. of Electrical and Electronic Eng., Niigata University, Niigata, Japan
E-mail: {shogo,schoi}@eng.niigata-u.ac.jp
† Graduate School of Science and Tech., Niigata University, Niigata, Japan
E-mail: takumi3143@telecom0.eng.niigata-u.ac.jp

***Abstract*—This paper proposes 3-D OCT data denoising with nonseparable oversampled lapped transform (NSOLT), and examines the effectiveness through experiments. NSOLT is a lattice-based redundant transform which simultaneously satisfies the symmetric, real-valued and compact-support property. It is possible to apply a dictionary learning technique to the design by preparing examples. NSOLT is capable of having rational redundancy by controlling the number of channels and decimation ratio. In this study, a denoising technique is proposed by combining learned NSOLT dictionary and iterative hard thresholding (IHT), and the performance of the proposed method is evaluated for 3-D OCT data. It is verified through robust median estimator of noise variance and structural similarity index measure (SSIM) that the proposed technique yields effective denoising performance with moderate redundancy.**

I. Introduction

Redundant transforms and sparse representation have successfully found a lot of applications in the field of image processing, such as feature extraction, denoising, deblurring, super-resolution, inpainting, as well as compressive sensing [1]–[6]. It is a natural demand to extend the sparsity-aware image processing for higher-dimensional signals such as volumetric images acquired through computed tomography (CT), magnetic resonance imaging (MRI) [7], and optical coherence tomography (OCT) [8]–[11].

An OCT system produces high-resolution cross-sectional images of microscopic structures in living tissue. A signal acquired by a 3-D OCT scanner is expressed as

$$S_{v,h}(d) = A_{v,h}(d) + \sum_{\hat{d}\in\mathcal{D}_{v,h}} \alpha(\hat{d}) B_{v,h}(d-\hat{d}) \cos\left(\frac{4\pi(d-\hat{d})}{\lambda}\right), \quad (1)$$

where v, h and d denote vertical, horizontal and depth position, respectively, and $\mathcal{D}_{v,h}$ and $\alpha(\hat{d})$ are a set of depth positions where tissues exist at spatial position (v, h) as coherent reflectors and an attenuation factor at $\hat{d}$. The first term, $A_{v,h}(d)$, in the right hand side of (1) denotes the mean (DC) intensity at position (v, h) and the second carries information about tissue structure, where λ is the center wavelength of the source light and $B_{v,h}(d-\hat{d})$ is a coherence function. The coherence function $B_{v,h}(d-\hat{d})$ shapes like Gaussian function and its peak indicates a tissue location.

What we are interested in is to obtain the second term, i.e., the target signal is

$$U_{v,h}(d) = S_{v,h}(d) - A_{v,h}(d) = \sum_{\hat{d}\in\mathcal{D}_{v,h}} \alpha(\hat{d}) B_{v,h}(d-\hat{d}) \cos\left(\frac{4\pi(d-\hat{d})}{\lambda}\right). \quad (2)$$

In practice, the acquired signal $S_{v,h}(d)$ is contaminated by noise and other degradation processes and $U_{v,h}(d)$ is required to be restored from the contaminated observation.

Let us assume that an observed OCT signal suffers from additive white Gaussian noise (AWGN) as

$$X_{v,h}(d) = U_{v,h}(d) + W_{v,h}(d), \quad (3)$$

where $X_{v,h}(d)$ is an observed coherence function and $W_{v,h}(d)$ is AWGN. This assumption is realistic because the DC intensity $A_{v,h}(d)$ occupies most energy of $S_{v,h}(d)$ and thus the sensor sensitivity severely affects to $U_{v,h}(d)$.

In this study, we propose to apply 3-D nonseparable oversampled lapped transforms (NSOLTs) [12] to OCT signal denoising, and experimentally evaluate the performance. NSOLT is an invertible redundant transform generated and implemented by a lattice structure [12]–[14]. The atoms are guaranteed to be symmetric or antisymmetric, and posses the nonseparable, real-valued, overlapping and compact-support property. NSOLT is also equipped with no-DC-leakage option and atom termination function at image boundary. Under a structural constraint of the Parseval tight frame property, we can also design an NSOLT dictionary with multiscale representation through a dictionary learning approach [12], [15]. The redundancy R can flexibly be controlled by the number of channels P and the downsampling ratio M.

In [12], it was verified that NSOLTs have comparable or superior sparse representation performance to the Sparse K-SVD, a learning-based dictionary developed by Rubinstein *et al.* [16], especially for pictures with large smooth region. In this work, we verify the significance of 3-D NSOLTs by applying it to a 3-D OCT data denoising problem and evaluating the performance.

II. Sparsity-Aware Denoising

In a sparsity-aware denoising approach, the signal modeling has a crucial role. Let $\mathbf{u} \in \mathbb{R}^N$ be a vectorized volume image

Data: Observed picture $\mathbf{x} \in \mathbb{R}^N$
Result: Denoised picture $\hat{\mathbf{u}} \in \mathbb{R}^M$
Initialization;
$i \leftarrow 0$;
$\mathbf{y}^{(0)} \leftarrow \mathbf{D}^T\mathbf{x}$;
Main iteration to find $\mathbf{y}$ that minimizes
$f(\mathbf{y}) = \|\mathbf{x} - \mathbf{D}\mathbf{y}\|_2^2$ s.t. $\|\mathbf{y}\|_0 \leq K$;
repeat
  $i \leftarrow i + 1$;
  $\mathbf{y}^{(i)} \leftarrow \mathcal{H}_{\lambda_K}\left(\mathbf{y}^{(i-1)} - \mu\mathbf{D}^T(\mathbf{D}\mathbf{y}^{(i-1)} - \mathbf{x})\right)$;
until $\|\mathbf{y}^{(i)} - \mathbf{y}^{(i-1)}\|_2^2 / \|\mathbf{y}^{(i)}\|_2^2 < \epsilon$;
$\hat{\mathbf{u}} \leftarrow \mathbf{D}\mathbf{y}^{(i)}$;

Algorithm 1: IHT algorithm with Parseval tight frame dictionary $\mathbf{D}$, where $\mathcal{H}_{\lambda_K}(\cdot)$ is a hard thresholding vector function defined by (6), $\mu = (1 - c)$ with $0 < c \ll 1$ and $\mathbf{D}^T$ is an adjoint of $\mathbf{D}$.

of $U_{v,h}(d)$ in (2). Then, vector $\mathbf{u}$ can be represented through a transform matrix $\mathbf{D} \in \mathbb{R}^{N \times L}$ as

$$\mathbf{u} = \mathbf{D}\mathbf{y} \tag{4}$$

with a coefficient vector $\mathbf{y} \in \mathbb{R}^L$, where $N < L$ for redundant $\mathbf{D}$. The column vectors of $\mathbf{D}$, denoted as $\{\mathbf{d}_\ell\}_{\ell=0}^{L-1}$, and $\mathbf{D}$ itself are referred to as 'atoms' and a 'dictionary,' respectively.

Sparse coding through some dictionary can approximate an image $\mathbf{x}$ and be used for removing AWGN from $\mathbf{x}$. This purpose is achieved by finding $\mathbf{y}$ of which nonzero coefficients are as few as possible under the constraint $\|\mathbf{x} - \mathbf{D}\mathbf{y}\|_2^2 \leq \epsilon$ for a small positive constant ϵ. The problem looking for an optimum $\mathbf{y}$ is formulated as follows:

$$\hat{\mathbf{y}} = \arg\min_{\mathbf{y} \in \mathbb{R}^L} \|\mathbf{x} - \mathbf{D}\mathbf{y}\|_2^2 \text{ s.t. } \|\mathbf{y}\|_0 \leq K, \tag{5}$$

where $\|\cdot\|_0$ denotes the count of nonzero coefficients and $0 < K \ll N$. Finally, an approximation or denoised vector of $\mathbf{x}$ is obtained by $\hat{\mathbf{u}} = \mathbf{D}\hat{\mathbf{y}}$.

The iterative hard thresholding (IHT) [17], [18] approximately solves (5). The advantage is the computational efficiency and it is applicable to huge amount of data. Algorithm 1 shows the IHT algorithm with a Parseval tight frame dictionary, where $\mathcal{H}_{\lambda_K}(\cdot)$ is a hard thresholding vector function defined by

$$\mathcal{H}_{\lambda_K}(\mathbf{v}) = \mathrm{sign}(\mathbf{v}) \circ |\mathbf{v}|_{(|\mathbf{v}| > \lambda_K)}, \tag{6}$$

where $\mathrm{sign}(\cdot)$ returns a sign vector which contains -1, 0 or $+1$ according to the sign of each element and $|\mathbf{v}|_{(|\mathbf{v}|>\lambda_K)}$ denotes a function that remains elements of which absolute value is greater than λ_K and replace the others to zero, where λ_K is the K-th largest absolute value among elements in $\mathbf{v}$. Symbol $\circ$ denotes element-wise product, i.e., Hadamard product.

III. 3-D NSOLT Dictionary

Developing a 3-D dictionary is a crucial task to improve the performance of sparsity-aware volume data processing.

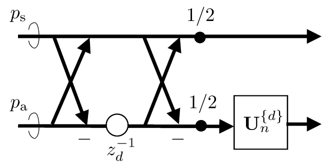


Fig. 1. Examples of lattice structures for realizing propagation matrix $\mathbf{V}_n^{\{i\}}\mathbf{Q}(z_i)$.

The nonseparable construction is indispensable so that slanting edges and textures are sparsely represented. In addition, redundancy R should be taken care of because the amount of data is huge and its influence to the required computational resources is not trivial.

Let us briefly review NSOLT as a dictionary $\mathbf{D}$ to model a volumetric image $\mathbf{u}$ as in (4). In the followings, $\mathbf{z} \in \mathbb{C}^3$ denotes a 3×1 complex variable vector $(z_0, z_1, z_2)^T$ in the 3-D z-transform domain, and P denotes the number of channels. Input arrays are assumed to be divided into small blocks according to a downsampling factor $\mathbf{M} \in \mathbb{Z}^{3\times 3}$, which determines the stride of atomic images. The downsampling ratio M and redundancy R are given by $M = |\det(\mathbf{M})|$ and $R = P/M(= L/N)$, respectively. Note that the product of sequential matrices $\mathbf{A}_N\mathbf{A}_{N-1}\cdots\mathbf{A}_1$ is denoted by $\prod_{n=1}^N \mathbf{A}_n$. Symbols $\mathbf{o}$, $\mathbf{O}$ and $\mathbf{I}_m$ are reserved for the null column vector, null matrix, and $m \times m$ identity matrix, respectively.

A. Lattice Structure of NSOLT

NSOLTs are composed only of symmetric or antisymmetric atoms [12]–[14]. According to the numbers of symmetric channels p_s and antisymmetric channels $p_a(= P - p_s)$, they are categorized into two types as follows:

- Type-I: $p_s = p_a$,
- Type-II: $p_s \neq p_a$.

In this study, we adopt only the Type-I construction, where the number of channels $P = p_s + p_a$ must be even. The polyphase matrix of 3-D analysis Type-I NSOLT is expressed by

$$\mathbf{E}(\mathbf{z}) = \prod_{i=0}^{2}\left\{\prod_{n=1}^{N_i}\left(\mathbf{V}_n^{\{i\}}\mathbf{Q}(z_i)\right)\right\} \cdot \mathbf{V}_0\mathbf{E}_0, \tag{7}$$

where $N_i \in (0 \cup \mathbb{N})$ is the polyphase order of the i-th dimension, which controls the overlap amount of atomic images, and

$$\mathbf{Q}(z_i) = \mathbf{B}_P^{(\frac{P}{2})}\begin{pmatrix}\mathbf{I}_{p_s} & \mathbf{O} \\ \mathbf{O} & z_i^{-1}\mathbf{I}_{p_a}\end{pmatrix}\mathbf{B}_P^{(\frac{P}{2})},$$

$$\mathbf{V}_n^{\{i\}} = \begin{pmatrix}\mathbf{I}_{p_s} & \mathbf{O} \\ \mathbf{O} & \mathbf{U}_n^{\{i\}}\end{pmatrix},$$

and

$$\mathbf{B}_P^{(m)} = \frac{1}{\sqrt{2}}\begin{pmatrix}\mathbf{I}_m & \mathbf{I}_m \\ \mathbf{I}_m & -\mathbf{I}_m\end{pmatrix}.$$

Fig. 1 illustrates a building block of the corresponding lattice structure, where $\mathbf{U}_n^{\{i\}} \in \mathbb{R}^{p_a \times p_a}$ is an arbitrary invertible

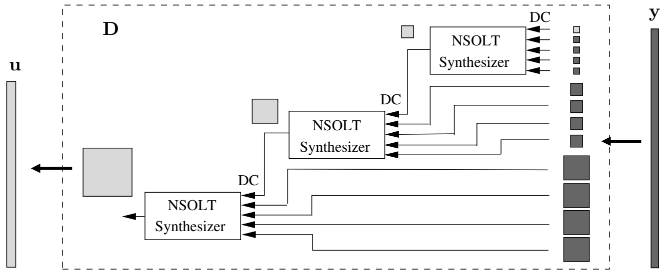


Fig. 2. Tree structure of multiscale NSOLT

matrix. For a diagonal decimation factor $\mathbf{M}$, we can adopt the product of 3-D separable discrete cosine transform (DCT) $\mathbf{E}_0 \in \mathbb{R}^{M\times M}$, which is obtained by a Kronecker product of three 1-D DCTs, and

$$\mathbf{V}_0 = \begin{pmatrix} \mathbf{W}_0 & \mathbf{O} \\ \mathbf{O} & \mathbf{U}_0 \end{pmatrix} \in \mathbb{R}^{P\times M}$$

as the initial matrix, where $\mathbf{W}_0 \in \mathbb{R}^{p_{\mathrm{s}}\times p_{\mathrm{s}}}$ and $\mathbf{U}_0 \in \mathbb{R}^{p_{\mathrm{a}}\times p_{\mathrm{a}}}$ are arbitrary left invertible parameter matrices.

B. Multiscale Tight Frame Representation

NSOLTs are equipped with the no-DC-leakage option and atom termination function as an image boundary manipulation [12], [13], [19]. In addition, when all of the parameter matrices $\mathbf{W}_0$, $\mathbf{U}_0$, and $\{\mathbf{U}_n^{\{i\}}\}$ are restricted to be orthonormal, it yields a Parseval tight frame.

A Parseval tight synthesis system is obtained as an adjoint of a Parseval tight analysis counterpart. In other words, we can obtain a paraunitary synthesis system as a paraconjugation of a paraunitary analysis counterpart. Furthermore, by iteratively reconstructing the DC coefficients as shown in Fig. 2, a multiscale tight frame dictionary is realized [12], [15]. The redundancy $R_M^P(\tau)$ of τ-level multiscale NSOLT is given by

$$R_M^P(\tau) = \begin{cases} (P-1)\tau + 1, & M = 1 \\ \frac{P-1}{M-1} - \frac{P-M}{(M-1)M^\tau}, & M \geq 2 \end{cases} . \quad (8)$$

C. Example-based Design

NSOLT can be designed through an example-based learning technique [12], [15]. The design problem is formulated as follows:

$$\{\hat{\mathbf{D}}, \{\hat{\mathbf{y}}_i\}\} = \arg\min_{\mathbf{D},\{\mathbf{y}_i\}} \sum_{i=0}^{S-1} \|\mathbf{x}_i - \mathbf{D}\mathbf{y}_i\|_2^2 \text{ s.t. } \|\mathbf{y}_i\|_0 \leq K, \quad (9)$$

where $\mathbf{x}_i$ is a training image and $\mathbf{y}_i$ is a coefficient vector of $\mathbf{x}_i$. As typical dictionary learning techniques, we divide the learning process into the 'sparse approximation' and 'dictionary update' stage, and alternatively apply these two stages to find a solution set.

- **Sparse approximation stage** tries to find a set of sparse coefficient vectors $\{\mathbf{y}_i\}$ which minimizes the approximation error for given images $\{\mathbf{x}_i\}$ under a fixed dictionary $\hat{\mathbf{D}}$ and the K-sparse constraint. This problem is formulated as follows:

$$\hat{\mathbf{y}}_i = \arg\min_{\mathbf{y}_i} \|\mathbf{x}_i - \hat{\mathbf{D}}\mathbf{y}_i\|_2^2 \text{ s.t } \|\mathbf{y}_i\|_0 \leq K$$

for $i = 0, 1, \cdots, S-1$, where S is the number of samples.
- **Dictionary update stage** tries to find a dictionary $\mathbf{D}$ which minimizes the approximation error for given images $\{\mathbf{x}_i\}$ and fixed coefficients $\{\hat{\mathbf{y}}_i\}$. Since a tight NSOLT can be designed by controlling rotation angles $\{\theta_i\}$ and sign parameters $\{s_i\}$ based on Givens factorization of orthonormal parameter matrices, the design problem is formulated as

$$\hat{\mathbf{\Theta}} = \arg\min_{\mathbf{\Theta}} \sum_{i=0}^{S-1} \|\mathbf{x}_i - \mathbf{D}_{\mathbf{\Theta}}\hat{\mathbf{y}}_i\|_2^2,$$

where $\mathbf{\Theta} = \{\{\theta_i\}, \{s_i\}\}$ and $\mathbf{D}_{\mathbf{\Theta}}$ is a multiscale NSOLT dictionary determined by $\mathbf{\Theta}$ [12], [15]. $\hat{\mathbf{D}} = \mathbf{D}_{\hat{\mathbf{\Theta}}}$ is used in next step.

IV. 3-D OCT Denoising with NSOLT

In this section, we analyze an acquired OCT signal and propose to apply 3-D NSOLT to denoising it for extracting the coherent reflection term. Our denoising procedure is summarized as follows:

1) Remove the DC intensity in the depth direction by subtracting the average slice of a current, ten previous and ten post slices from every slice. Circular convolution is adopted to maintain the boundaries in the light direction.
2) Draw a small volume patch $\mathbf{x}_0$ that contains significant intensity variation from the observed volume $\mathbf{x}$ that is obtained after DC subtraction in the previous step.
3) Design a 3-D NSOLT dictionary $\mathbf{D}$ by using $\mathbf{x}_0$ as an example through the dictionary learning approach described in III-C.
4) Denoise $\mathbf{x}$ by applying IHT and the designed 3-D NSOLT.

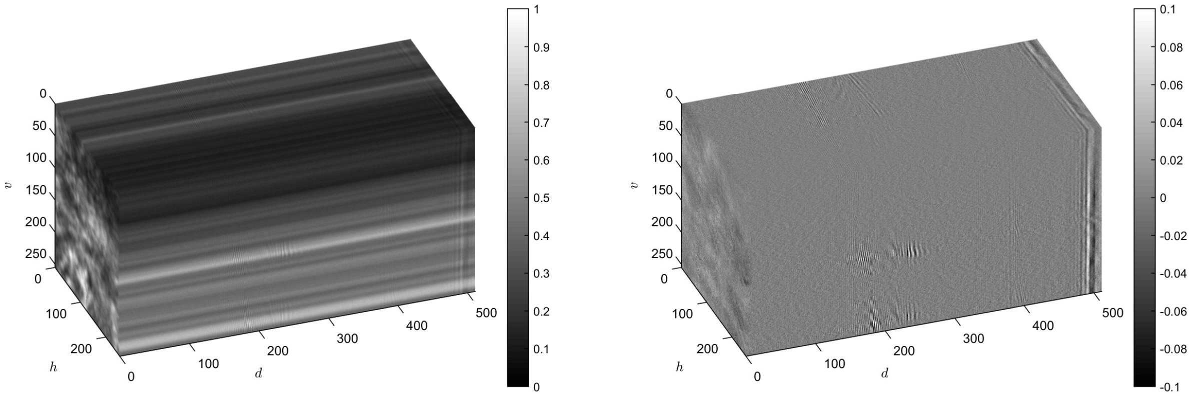
The reason why we take an average of succeeding twenty-one slices in Step 1 is because we observed strong frequency components in the original OCT signal $S_{v,h}(d)$ lower than around $\pi/20$ from the Fourier analysis in the depth direction.

A. 3-D OCT Image

This work uses the 3-D OCT data shown in Fig. 3, which was acquired by multifrequency sweeping Fizeau interferometer [9]–[11]. Kidney tissue of a mouse fixed with formalin and embedded in paraffin was used as the test object.

Fig. 3 (a) shows the original volumetric image of size $256\times 256\times 512$ voxels in unsigned 16bpp grayscale and (b) shows the DC-subtracted one, which we process as an observation $\mathbf{x}$. We see that the coherent reflections in $\mathbf{x}$ are not so strong that noise intensity becomes relatively high.

Signal $X_{v,h}(d)$ in Fig. 4 is a sequence of the volumetric image shown in Fig. 3 (b) at the vertical and horizontal center along the light direction. A strong intensity is observed at slice 183. This implies that some object exits around the



(a) (b)

Fig. 3. 3-D OCT images of size $256 \times 256 \times 512$. (a) Raw 3-D OCT data in unsigned 16bpp grayscale, where the intensity is normalized to the range $[0, 1]$. (b) DC-subtracted OCT data $X_{v,h}(d)$ in signed 16bpp grayscale.

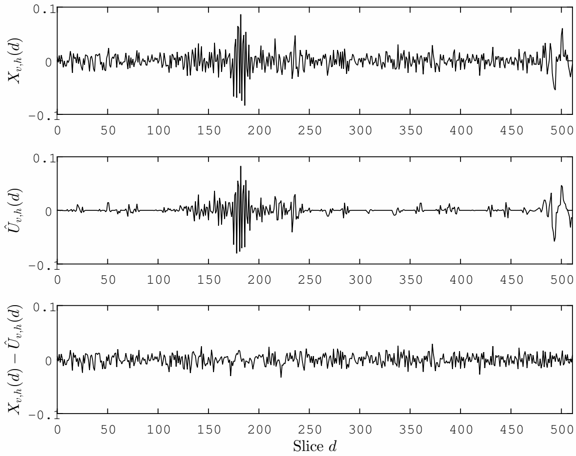


Fig. 4. Intensity sequences of 3-D OCT data in Fig. 3 in the light direction at vertical and horizontal center position.

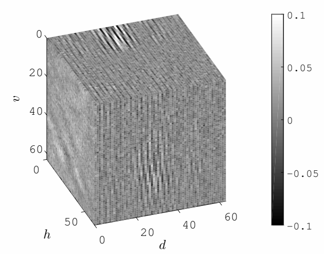


Fig. 5. Volume patch $\mathbf{x}_0$ of size $64 \times 64 \times 64$ drawn from Fig. 3 (b) for 3-D NSOLT dictionary learning.

slice and detected by OCT as a coherent reflection. From this observation, we drew a volume patch $\mathbf{x}_0$ of size $64 \times 64 \times 64$ voxels around the peak location for the 3-D NSOLT dictionary learning. Fig. 5 shows the extracted volume patch.

TABLE I
EXPERIMENTAL SETTINGS OF 3-D MULTISCALE NSOLT LEARNING.

♯Downsampl. M	$1 \times 1 \times 2$
♯Channels P	$3 + 3$
Polyphase Ord. N_d	$(2, 2, 2)$
Sparsity K/N	$32,768/64^3$
Initial dictionary	Haar DWT
Training data size	$64 \times 64 \times 64$
Sparse Approx.	IHT
Iterations	20

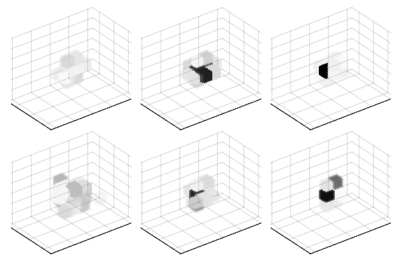


Fig. 6. Learned atoms of size $3 \times 3 \times 6$ with $\mathbf{M} = \mathrm{diag}(M_0, M_1, M_2) = \mathrm{diag}(1, 1, 2)$, $M = |\det(\mathbf{M})| = M_0 \times M_1 \times M_2 = 1 \times 1 \times 2$, $P = p_\mathrm{s} + p_\mathrm{a} = 3 + 3$, $(N_0, N_1, N_2)^T = (2, 2, 2)^T$ and $\tau = 3$, where single level atoms are shown.

B. 3-D NSOLT Dictionary Learning

We trained 3-D NSOLT dictionaries by using the volume patch $\mathbf{x}_0$ shown in Fig. 5, where a single example is adopted, i.e., $S = 1$. Since the patch size is larger than atoms in an NSOLT dictionary, this example also has similar effects to the case of using multiple small examples.

The experimental settings are summarized in Table I. In this work, the no-DC-leakage property is taken into account and the atom termination at image boundary is adopted. We examined systems with $P = p_\mathrm{s} + p_\mathrm{a} = 3 + 3$ for different tree levels. In the dictionary update stage, the unconstrained minimization function '`fminunc`' of MATLAB R2015a is

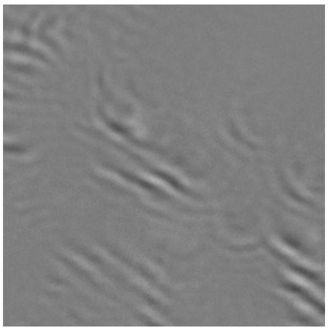


Fig. 7. The v-h slice of the volume image shown in Fig. 3 (b) at $d = 183$.

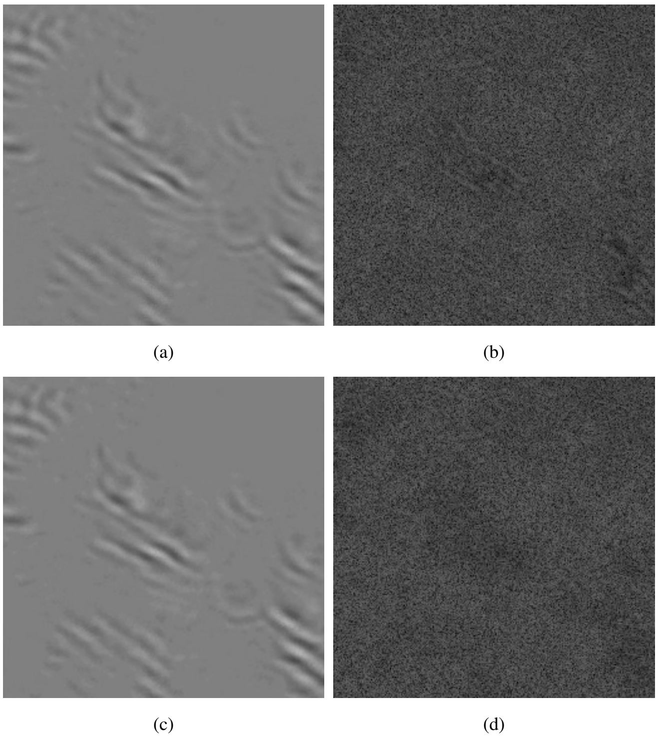


Fig. 8. Examples of v-h slices at $d = 183$. (a) Denoised slice with three-level NSOLT in Fig. 6, (b) Absolute difference between (a) and Fig. 7, (c) Denoised slice with single-level undecimated Haar transform, (d) Absolute difference between (c) and Fig. 7.

used for optimizing angles $\{\theta_i\}$ after initialization of the signs $\{s_i\}$ and angles $\{\theta_i\}$ with the genetic algorithm function 'ga' of MATLAB R2015a. Fig. 6 shows an example set of learned atoms.

V. Denoising Performance Evaluation

Let us evaluate the denoising performance of the proposed method for the volume data shown in Fig. 3 (b). In this experiment, we adopt the IHT algorithm with Parseval tight 3-D NSOLTs, where the sparsity is set to remain $K = 2,097,152$ coefficients in total, i.e., 6.25% of the number of the spacial voxels $256 \times 256 \times 512$, the number of channels is fixed to $P = p_{\mathrm{s}} + p_{\mathrm{a}} = 3 + 3$ and the tree level τ is set to three.

Sequence "$\hat{U}_{v,h}(d)$" in Fig. 4 shows the intensity variation of the denoised volume data in the light direction at the vertical and horizontal center location. Sequence "$X_{v,h}(d) - \hat{U}_{v,h}(d)$" denotes the difference between the observed sequence $X_{v,h}(d)$ and the denoised one. It is observed that noisy fluctuations are suppressed compared with the original while maintaining the significant coherent reflections. The peaks of such significant coherent reflections mean the positions where kidney tissues exist and convey the structural information. The proposed method is expected to avoid false-positive detections. Fig. 7 shows the v-h slice of the volume image shown in Fig. 3 (b) at $d = 183$, and Figs. 8 (a) and (b) show the v-h slices of the denoised volume data and the absolute difference from the original, respectively. For reference, those of the undecimated Haar transform (UHT) [3], [20] are also given in Figs. 8 (c) and (d). Note that UHT is Parseval tight and is applicable to the IHT algorithm shown in 1. From Fig. 8, we can see that 3-D NSOLT appropriately removes the noisy component while keeping the coherent pattern. Note that the redundancy of the three-level NSOLT is only 4.5, while that of the single-level UHT is 8.

TABLE II

Denoising results of IHT with 3-D NSOLT for volume data of size $256 \times 256 \times 512$, where $K = 2,097,152$. "σ^2_{RME}" denotes robust median estimation of noise variance.

Redundancy R		< 5	
$\sharp$Downsampl. M		$1 \times 1 \times 2$	
$\sharp$Channels P		$p_s + p_a = 3 + 3$	
Evaluation		σ^2_{RME}	SSIM ($d = 183$)
Tree level	$\tau = 1$	1.03×10^{-6}	0.9455
	$\tau = 2$	1.68×10^{-6}	0.9440
	$\tau = 3$	3.60×10^{-7}	0.9418
	$\tau = 4$	5.90×10^{-7}	0.9413
	$\tau = 5$	4.95×10^{-7}	0.9420

TABLE III

Denoising result of IHT with single-level undecimated Haar transform for volume data of size $256 \times 256 \times 512$, where $K = 2,097,152$. "σ^2_{RME}" denotes robust median estimation of noise variance.

Redundancy R		$7\tau + 1$	
$\sharp$Downsampl. M		$1 \times 1 \times 1$	
$\sharp$Channels P		$p_s + p_a = 4 + 4$	
Evaluation		σ^2_{RME}	SSIM ($d = 183$)
Tree level	$\tau = 1$	0.00	0.9329

Table II summarizes the robust median estimation (RME) of noise variances [21] in the full volume data and structural similarity index measure (SSIM) [22] of slices at $d = 183$. The former and latter measure the ability of noise removal and structure preservation characteristic, respectively. Note here that we extract fine components from signals by adopting the 3-D high-pass filter with the following impulse response:

$$H_0 = \frac{1}{2\sqrt{2}} \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}, H_1 = \frac{1}{2\sqrt{2}} \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}$$

for calculating RME of noise variance, where H_d denotes the impulse response of the d-th slice.

For reference, we also give the denoising results with the single-level UHT in Table III. Although the estimated noise variance is larger than that of UHT, 3-D NSOLT seems to preserve the structure of more significant coherent reflections than UHT. Unfortunately, we are not allowed to access the noiseless clean signal and SSIM does not give appropriate evaluation in this experiment. We can subjectively see that 3-D NSOLT shows comparable performance to UHT with less redundancy in the IHT denoising approach. One of the reasons is that 3-D NSOLT can be designed through the example-based approach shown in III-C and efficiently model a given volumetric image.

VI. Conclusions

In this work, we porpoised to apply IHT and 3-D NSOLT to OCT data denoising. By using dictionary learning approach, we obtained appropriate atom set for sparsely representing coherent reflections. Through some experiments, we verified the combination of IHT and NSOLT shows comparable or superior performance to the combination of IHT and the undecimated Haar transform with significant reduction of the redundancy.

Acknowledgment

This work was supported by JSPS KAKENHI under Grant 26420347. The authors would like to thank the anonymous reviewers for their valuable comments and suggestions to improve the quality of the paper.

References

[1] S. Mallat, *A Wavelet Tour of Signal Processing, Third Edition: The Sparse Way*, Academic Press, 2008.

[2] D. P. Palomar and Y. C. Eldar, Eds., *Convex Optimization in Signal Processing and Communications*, Cambridge University Press, 2009.

[3] J.-L. Starck, F. Murtagh, and J. M. Fadili, *Sparse Image and Signal Processing: Wavelets, Curvelets, Morphological Diversity*, Cambridge University Press, 2010.

[4] M. Elad, *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*, Springer, 2010.

[5] Y. C. Eldar and G. Kutyniok, Eds., *Compressed Sensing, Theory and Applications*, Cambridge University Press, 2012.

[6] M. Elad, M. A. T. Figueiredo, and Y. Ma, "On the role of sparse and redundant representations in image processing," *Proc. IEEE*, vol. 98, no. 6, pp. 972–982, June 2010.

[7] N. Nikolaidis and I. Pitas, *3-D Image Processing Algorithms*, Wiley-Interscience, 2001.

[8] J. M. Schmitt, "Optical coherence tomography (OCT): A review," *IEEE J. Sel. Topics Quantum Electron.*, vol. 5, no. 4, July/Aug. 1999.

[9] S. Choi, T. Watanabe, O. Sasaki, and T. Suzuki, "OCT based on multi-frequency sweeping fizeau interferometer with phase modulating method," in *Proc. SPIE 8839, Dimensional Optical Metrology and Inspection for Practical Applications II*, Sept. 2013, 88380D.

[10] S. Choi, J. Hukumoto, O. Sasaki, and T. Suzuki, "Multi-frequency sweeping sinusoidal phase modulated fizeau interferometer for OCT," in *Proc. of 17th Opto-Electronics and Communications Conference (OECC)*, 2012, pp. 568–569.

[11] S. Choi, T. Watanabe, T. Suzuki, F. Nin, H. Hibino, and O. Sasaki, "Multifrequency swept common-path en-foce OCT for wide-field measurement of interior surface vibrations in thick biological tissues," *Opt Express*, vol. 23, no. 16, pp. 21078–21089, Aug. 2015.

[12] S. Muramatsu, K. Furuya, and N. Yuki, "3-D NSOLT for volume data sparse representation," in *Proc. of IEICE SIP Symp.*, Nov. 2014, pp. 1–6, in Japanese.

[13] S. Muramatsu and N. Aizawa, "Lattice structures for 2-D non-separable oversampled lapped transforms," in *Proc. of IEEE ICASSP*, May 2013, pp. 5632–5636.

[14] S. Muramatsu and N. Aizawa, "Image restoration with 2-D non-separable oversampled lapped transforms," in *Proc. of IEEE ICIP*, Sept. 2013, pp. 1051–1055.

[15] S. Muramatsu, "Structured dictionary learning with 2-D non-separable oversampled lapped transform," in *Proc. of IEEE ICASSP*, May 2014, pp. 2643–2647.

[16] R. Rubinstein, M. Zibulevsky, and M. Elad, "Double sparsity: Learning sparse dictionaries for sparse signal approximation," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1553–1564, Mar. 2010.

[17] T. Blumensath and M. E. Davies, "Iterative hard thresholding for compressed sensing," *Appl. Comput. Harmon. Anal.*, vol. 27, no. 3, pp. 265–274, Nov. 2009.

[18] T. Blumensath and M. E. Davies, "Normalized iterative hard thresholding: Guaranteed stability and performance," *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 2, pp. 298–309, Apr. 2010.

[19] K. Furuya, S. Hara, and S. Muramatsu, "Boundary operation of 2-D non-separable oversampled lapped transforms," in *Proc. of APSIPA ASC*, Nov. 2013.

[20] M. Unser, "Texture classification and segmentation using wavelet frames," *IEEE Trans. Image Process.*, vol. 4, no. 11, pp. 1549–1560, Nov. 1995.

[21] S. G. Chang, B. Yu, and M. Vetterli, "Adaptive wavelet thresholding for image denoising and compression," *IEEE Trans. Image Process.*, vol. 9, no. 9, pp. 1532–1546, Sept. 2000.

[22] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.