STRUCTURED DICTIONARY LEARNING WITH 2-D NON-SEPARABLE OVERSAMPLED LAPPED TRANSFORM

*Shogo MURAMATSU**

Niigata University, Dept. of Electrical and Electronic Eng.
8050 2-no-cho Ikarashi, Nishi-ku, Niigata, 950-2181, JAPAN
shogo@eng.niigata-u.ac.jp

ABSTRACT

This work proposes a novel design method of a two-dimensional (2-D) Non-Separable Oversampled Lapped Transform (NSOLT) for a given image by introducing a typical two stage procedure of dictionary learning. NSOLT is a lattice-structure-based transform and yields a redundant dictionary of which atoms satisfy the non-separable, symmetric, real-valued, overlapping and compact-support property. In addition, the Parseval tight frame constraint can structurally be imposed, while the redundancy $\mathcal{R}$ is flexibly controlled by the ratio of the number of channels P and the downsampling ratio M. Compared with the other dictionary learning approaches, the proposed method is moderately structured so that it is capable of multiscale construction as well as atom termination at image boundary. The significance of the proposed method is verified by showing an example of learned dictionary and sparse approximation results.

Index Terms— Dictionary learning, Parseval tight frame, Multiscale representation, Iterative hard thresholding, NSOLT

1. INTRODUCTION

Redundant transforms are indispensable for sparsely representing images. Combined with state-of-the-art optimization techniques, redundant systems have found a lot of image processing applications, e.g. feature extraction, denoising, deblurring, super-resolution, inpainting, as well as compressive sensing [1–6].

Let $\mathbf{x} \in \mathbb{R}^N$ be a vectorized image. Then, a redundant transform $\mathbf{D} \in \mathbb{R}^{N\times L}$ can be used to represent $\mathbf{x}$ as

$$\mathbf{x} = \mathbf{D}\mathbf{y} \tag{1}$$

with a coefficient vector $\mathbf{y} \in \mathbb{R}^L$, where $N < L$. That is, $\mathbf{x}$ is represented as a linear-combination of column vectors $\{\mathbf{d}_\ell\}_{\ell=0}^{L-1}$ of $\mathbf{D}$ with coefficients $\{y_\ell\}_{\ell=0}^{L-1}$ of $\mathbf{y}$ as weights. There exist infinite number of candidates for vector $\mathbf{y}$ when $\mathrm{rank}(\mathbf{D}) = N$. Vectors $\{\mathbf{d}_\ell\}_{\ell=0}^{L-1}$ and transform $\mathbf{D}$ are referred to as 'atoms' as primitive codewords of images and a 'dictionary' as a set of atoms, respectively. The purpose of sparse representation is to compactly represent or approximate a given image $\mathbf{x}$. For a given dictionary $\mathbf{D}$, this purpose is achieved by finding $\mathbf{y}$ of which nonzero coefficients are as few as possible under the constraint (1) or $\|\mathbf{x} - \mathbf{D}\mathbf{y}\|_2^2 \leq \epsilon$ for a small positive constant ϵ. If (1) is satisfied with K coefficients of $\mathbf{y}$, $\mathbf{x}$ is said to be K-sparse over dictionary $\mathbf{D}$, where we are concerned with the case $K \ll N$. When $\mathbf{x}$ is known to be exactly or approximately K-sparse over $\mathbf{D}$, the problem looking for an optimum $\mathbf{y}$ is formulated as follows:

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

where $\|\cdot\|_0$ denotes the count of nonzero coefficients. Unfortunately, solving the sparsity problem in (2) is known to be NP-hard. Thus, numerous algorithms have been proposed to approximately solve (2). Orthogonal matching pursuit (OMP) [7, 8], gradient pursuit (GP) [9] and iterative hard thresholding (IHT) [10, 11] are good examples of such algorithms.

The choice of the dictionary $\mathbf{D}$ is as important as that of sparse approximation algorithm since it determines the model of given images and severely influences the sparse representation performance. From innumerable candidates of dictionaries, we have to adopt one according to the target images and application. The dictionaries are categorized into two types: the analytical type and the learning-based type [12]. An analytical dictionary is mathematically predetermined and highly structured. Examples in this type of dictionaries include the shift-invariant Haar wavelets [13], Curvelets [3] and Contourlets [14, 15]. In general, the analytical dictionaries require less computational resources in its implementation than the latter type. On the other hand, learning-based dictionaries have an advantage that they can be finely tuned to given images. Several approaches to train a dictionary have been proposed so far [12, 16]. They include MOD [17], K-SVD [18], and SimCO [19].

Typical learning-based approaches adopt explicit matrix representation. Therefore, dictionaries generated by such approaches are likely to require more computational resources than analytical ones due to the unstructured nature. In order to resolve the computational complexity, structured dictionary learning approaches have also been studied. One successful example is Sparse K-SVD [20], which factorizes the dictionary $\mathbf{D}$ into a predetermined base dictionary $\mathbf{\Phi}$ and sparse matrix $\mathbf{A}$, i.e. $\mathbf{D} = \mathbf{\Phi}\mathbf{A}$. The reduction of memory requirements and implementation complexity is quite beneficial in practical applications especially for high-dimensional signals.

If a dictionary $\mathbf{D}$ satisfies $\mathbf{D}\mathbf{D}^T = B\mathbf{I}$ for some positive constant B, the set of the atoms $\{\mathbf{d}_\ell\}_{\ell=0}^{L-1}$ is said to be a tight frame for $\mathbb{R}^N$, where $\mathbf{I}$ is the identity matrix. Furthermore, when $B = 1$, the set is referred to as '1-tight frame' or 'Parseval tight frame.' In this case, the Parseval's equality $\|\mathbf{x}\|_2^2 = \|\mathbf{D}^T\mathbf{x}\|_2^2$ holds. That is, the energy of image $\mathbf{x}$ is preserved by the coefficient vector $\mathbf{y}^* = \mathbf{D}^T\mathbf{x}$. Such frames have an advantage that they make some sparsity-aware algorithms tractable and simpler [12]. For example, a sufficient condition of the convergence of IHT is simply met by setting the step size a constant less than unit [11].

It is not trivial to impose the tight frame constraint onto learning-based approaches with explicit matrix representation. In this study,

*This work was supported by JSPS KAKENHI (23560443).

978-1-4799-2893-4/14/$31.00 ©2014 IEEE

we introduce a 2-D non-separable oversampled lapped transform (NSOLT) [21, 22] as the constraint. NSOLT is an invertible redundant transform generated and implemented by a lattice structure, which is a generalized version of a 2-D non-separable lapped orthogonal transform [23–27], where the atoms are guaranteed to be symmetric or antisymmetric, and posses the non-separable, real-valued, overlapping and compact-support property. NSOLT is also equipped with no-DC-leakage option [21] and atom termination function at image boundary [28]. The Parseval tight frame constraint can structurally be imposed, while the redundancy $\mathcal{R}$ is flexibly controlled. In addition, multiscale representation is available. In order to show the significance of the proposed method, a design example is shown and the sparse approximation results under different settings are compared with those of Sparse K-SVD.

2. REVIEW OF 2-D NSOLT

Let us briefly review NSOLT as a dictionary $\mathbf{D}$ to model an image $\mathbf{x}$ as in (1) [21]. In the followings, $\mathbf{z} \in \mathbb{C}^2$ denotes a 2×1 complex variable vector $(z_\mathrm{y}, z_\mathrm{x})^T$ in the 2-D z-transform domain, and P denotes the number of channels. The coefficients are assumed to be downsampled by a factor $\mathbf{M} \in \mathbb{Z}^{2\times 2}$. The downsampling ratio M and redundancy $\mathcal{R}$ are given by $M = |\det(\mathbf{M})|$ and $\mathcal{R} = P/M(= L/N)$, respectively. As a special case, $\mathbf{M}$ can be the identity so that a shift-invariant system is generated.

2.1. Two types of lattice structure

NSOLTs are categorized into two types according to the number of symmetric channels p_s and the number of antisymmetric channels $p_\mathrm{a}(= P - p_\mathrm{s})$. Let us give an overview of these two types, where we consider only the case that the vertical polyphase order N_y and horizontal one N_x are all even. As a utility notation of butterfly matrices, we introduce the following matrix:

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

where $m = \min(p_\mathrm{s}, p_\mathrm{a})$ and $\mathbf{O}$ is a null matrix.

2.1.1. *Type-I NSOLT* ($p_\mathrm{s} = p_\mathrm{a}$)

When the number of channels P is even, it is possible to set $p_\mathrm{s} = p_\mathrm{a} = P/2$. From [21, Theorem 3], a Type-I lattice of NSOLT can be constructed as shown in Fig. 1 (a). The corresponding analysis polyphase matrix $\mathbf{E}(\mathbf{z})$ [29] is represented by

$$\mathbf{E}(\mathbf{z}) = \prod_{n_\mathrm{y}=1}^{N_\mathrm{y}} \left\{ \mathbf{R}_{n_\mathrm{y}}^{\{\mathrm{y}\}} \mathbf{Q}(z_\mathrm{y}) \right\} \cdot \prod_{n_\mathrm{x}=1}^{N_\mathrm{x}} \left\{ \mathbf{R}_{n_\mathrm{x}}^{\{\mathrm{x}\}} \mathbf{Q}(z_\mathrm{x}) \right\} \cdot \mathbf{R}_0 \mathbf{E}_0, \quad (3)$$

where $\mathbf{E}(\mathbf{z})$ is of size $P \times M$ and

$$\mathbf{Q}(z_d) = \mathbf{B}_P^{(\frac{P}{2})} \begin{pmatrix} \mathbf{I}_{p_\mathrm{s}} & \mathbf{O} \\ \mathbf{O} & z_d^{-1}\mathbf{I}_{p_\mathrm{a}} \end{pmatrix} \mathbf{B}_P^{(\frac{P}{2})}, \ \mathbf{R}_n^{\{d\}} = \begin{pmatrix} \mathbf{I}_{p_\mathrm{s}} & \mathbf{O} \\ \mathbf{O} & \mathbf{U}_n^{\{d\}} \end{pmatrix}.$$

In the above expressions, $\mathbf{I}_p$ denotes the identity matrix of size $p \times p$ and $\mathbf{U}_n^{\{d\}} \in \mathbb{R}^{p_\mathrm{a} \times p_\mathrm{a}}$ is an arbitrary nonsingular matrix. We adopt the 2-D discrete cosine transform (DCT) as $\mathbf{E}_0 \in \mathbb{R}^{M \times M}$ and

$$\mathbf{R}_0 = \begin{pmatrix} \mathbf{W}_0 & \mathbf{O} \\ \mathbf{O} & \mathbf{U}_0 \end{pmatrix} \begin{pmatrix} \mathbf{I}_{\lceil M/2 \rceil} & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \\ \mathbf{O} & \mathbf{I}_{\lfloor M/2 \rfloor} \\ \mathbf{O} & \mathbf{O} \end{pmatrix} \in \mathbb{R}^{P \times M},$$

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 nonsingular matrices.

2.1.2. *Type-II NSOLT* ($p_\mathrm{s} \neq p_\mathrm{a}$)

When $p_\mathrm{s} \neq p_\mathrm{a}$, we have to use a Type-II lattice structure [21, Theorem 3]. Fig. 1 (b) shows an analysis lattice structure of Type-II NSOLT. For $p_\mathrm{s} > p_\mathrm{a}$, the polyphase matrix $\mathbf{E}(\mathbf{z})$ is represented by

$$\mathbf{E}(\mathbf{z}) = \prod_{\ell_\mathrm{y}=1}^{N_\mathrm{y}/2} \left\{ \mathbf{R}_{\mathrm{E}\ell_\mathrm{y}}^{\{\mathrm{y}\}} \mathbf{Q}_\mathrm{E}(z_\mathrm{y}) \mathbf{R}_{\mathrm{O}\ell_\mathrm{y}}^{\{\mathrm{y}\}} \mathbf{Q}_\mathrm{O}(z_\mathrm{y}) \right\} \times \prod_{\ell_\mathrm{x}=1}^{N_\mathrm{x}/2} \left\{ \mathbf{R}_{\mathrm{E}\ell_\mathrm{x}}^{\{\mathrm{x}\}} \mathbf{Q}_\mathrm{E}(z_\mathrm{x}) \mathbf{R}_{\mathrm{O}\ell_\mathrm{x}}^{\{\mathrm{x}\}} \mathbf{Q}_\mathrm{O}(z_\mathrm{x}) \right\} \cdot \mathbf{R}_0 \mathbf{E}_0, \quad (4)$$

where $\mathbf{E}(\mathbf{z})$ is of size $P \times M$,

$$\mathbf{Q}_\mathrm{E}(z_d) = \mathbf{B}_P^{(p_\mathrm{a})} \begin{pmatrix} \mathbf{I}_{P-p_\mathrm{a}} & \mathbf{O} \\ \mathbf{O} & z_d^{-1}\mathbf{I}_{p_\mathrm{a}} \end{pmatrix} \mathbf{B}_P^{(p_\mathrm{a})}, \ \mathbf{R}_{\mathrm{E}\ell}^{\{\mathrm{y}\}} = \begin{pmatrix} \mathbf{W}_\ell^{\{d\}} & \mathbf{O} \\ \mathbf{O} & \mathbf{I}_{p_\mathrm{a}} \end{pmatrix},$$

$$\mathbf{Q}_\mathrm{O}(z_d) = \mathbf{B}_P^{(p_\mathrm{a})} \begin{pmatrix} \mathbf{I}_{p_\mathrm{a}} & \mathbf{O} \\ \mathbf{O} & z_d^{-1}\mathbf{I}_{P-p_\mathrm{a}} \end{pmatrix} \mathbf{B}_P^{(p_\mathrm{a})}, \ \mathbf{R}_{\mathrm{O}\ell}^{\{\mathrm{x}\}} = \begin{pmatrix} \mathbf{I}_{p_\mathrm{s}} & \mathbf{O} \\ \mathbf{O} & \mathbf{U}_\ell^{\{d\}} \end{pmatrix}.$$

In the above expressions, $\mathbf{W}_\ell^{\{d\}} \in \mathbb{R}^{p_\mathrm{s} \times p_\mathrm{s}}$ and $\mathbf{U}_\ell^{\{d\}} \in \mathbb{R}^{p_\mathrm{a} \times p_\mathrm{a}}$ are arbitrary nonsingular matrices. $\mathbf{E}_0 \in \mathbb{R}^{M \times M}$ is the 2-D DCT and

$$\mathbf{R}_0 = \begin{pmatrix} \mathbf{W}_0 & \mathbf{O} \\ \mathbf{O} & \mathbf{U}_0 \end{pmatrix} \begin{pmatrix} \mathbf{I}_{\lceil M/2 \rceil} & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \\ \mathbf{O} & \mathbf{I}_{\lfloor M/2 \rfloor} \\ \mathbf{O} & \mathbf{O} \end{pmatrix} \in \mathbb{R}^{P \times M},$$

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 nonsingular matrices.

2.2. Parseval tight frame constraint

If all of the parameter matrices $\{\mathbf{W}_\ell^{\{d\}}\}$, $\{\mathbf{U}_\ell^{\{d\}}\}$, $\mathbf{W}_0$ and $\mathbf{U}_0$ are constrained to be orthonormal, then $\mathbf{E}(\mathbf{z})$ becomes paraunitary. Any orthonormal matrix of size $p \times p$ can be factorized into a combination of $\frac{p(p-1)}{2}$ Givens rotations with a $p \times p$ diagonal matrix of which p diagonal elements take either of '$+1$' or '-1' [30]. By using this factorization, we can structurally constrain the parameter matrices to be orthonormal. In summary, the paraunitary property of $\mathbf{E}(\mathbf{z})$ can be maintained structurally. Table 1 summarizes the required number of parameters in this approach of paraunitary system construction.

From the frame-theoretic point of view, a paraunitary system corresponds to a tight frame [31]. Since the frame bound $B = 1$, NSOLT yields a 1-tight frame, i.e. Parseval tight frame. The paraconjugation of a paraunitary analysis system is known to yield a paraunitary synthesis system. In the real coefficient case, the paraconjugation of $\mathbf{E}(\mathbf{z})$ is represented by

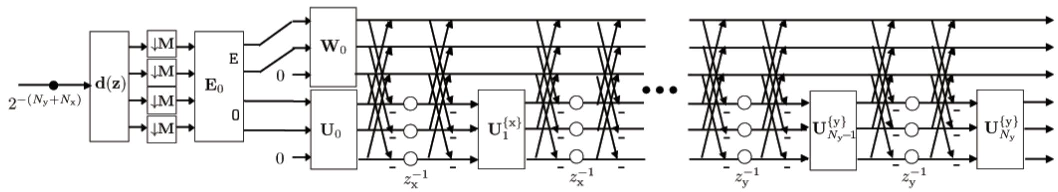
$$\mathbf{R}(\mathbf{z}) = \mathbf{z}^{-\bar{\mathbf{n}}} \mathbf{E}^T(\mathbf{z}^{-\mathbf{I}}), \quad (5)$$

where $\bar{\mathbf{n}} = [N_\mathrm{y}, N_\mathrm{x}]^T$ and $\mathbf{I}$ is the identity matrix of size 2×2.

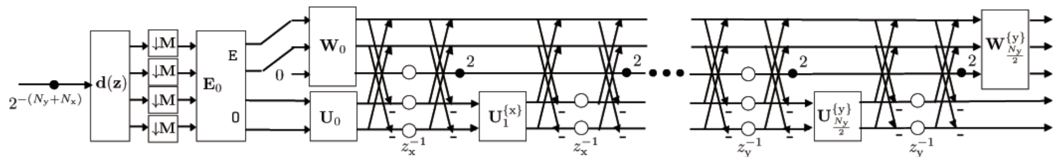
In the redundant case, i.e. $\mathcal{R} > 1$, there is an infinite number of perfect reconstruction (PR) combination of analysis and synthesis systems [32, 33]. It can be verified that the above pair of systems constitute a PR system together. The synthesis process with $\mathbf{R}(\mathbf{z})$ corresponds to a linear operation with a dictionary $\mathbf{D}$, and the analysis process with $\mathbf{E}(\mathbf{z})$ corresponds to the adjoint operation with $\mathbf{D}^T$.

2.3. Multiscale representation

At least, one channel in a PR analysis and synthesis system should take charge of (piecewise) DC component. If only one channel takes the role, the system is said to have no DC-leakage since the other channels annihilate the DC component [30]. The no-DC-leakage



(a) Type-I lattice, where $M = 4$, $P = p_\mathrm{s} + p_\mathrm{a} = 3 + 3 = 6$ and $\mathcal{R} = P/M = 3/2$.



(b) Type-II lattice, where $M = 4$, $P = p_s + p_a = 3 + 2 = 5$ ane $\mathcal{R} = P/M = 5/4$.

Fig. 1. Examples of 2-D NSOLT analysis lattice structures, where $\mathbf{d}(\mathbf{z})$ denotes the shifter chain determined by the downsampling factor $\mathbf{M}$. The polyphase orders N_x and N_y are assumed to be even.

Table 1. The number of design parameters for paraunitary NSOLT, where the polyphase orders N_x and N_y are assumed to be even.

Type	Param.	$\mathbf{W}_0$	$\mathbf{U}_0$	$\{\mathbf{W}_\ell^{\{d\}}\}$	$\{\mathbf{U}_\ell^{\{d\}}\}$	Total
I	$\sharp\{\theta_i\}$	$p_\mathrm{s}(p_\mathrm{s}-1)/2$	$p_\mathrm{a}(p_\mathrm{a}-1)/2$	-	$(N_\mathrm{x}+N_\mathrm{y})p_\mathrm{a}(p_\mathrm{a}-1)/2$	$(N_\mathrm{x}+N_\mathrm{y}+2)P(P-2)/8$
	$\sharp\{s_i\}$	p_s	p_a	-	$(N_\mathrm{y}+N_\mathrm{x})p_\mathrm{a}$	$(N_\mathrm{x}+N_\mathrm{y}+2)P/2$
II	$\sharp\{\theta_i\}$	$p_\mathrm{s}(p_\mathrm{s}-1)/2$	$p_\mathrm{ta}(p_\mathrm{a}-1)/2$	$(N_\mathrm{x}+N_\mathrm{y})p_\mathrm{s}(p_\mathrm{s}-1)/4$	$(N_\mathrm{x}+N_\mathrm{y})p_\mathrm{a}(p_\mathrm{a}-1)/4$	$(N_\mathrm{x}+N_\mathrm{y}+2)(p_\mathrm{s}^2+p_\mathrm{a}^2-P)/4$
	$\sharp\{s_i\}$	p_s	p_a	$(N_\mathrm{x}+N_\mathrm{y})p_\mathrm{s}/2$	$(N_\mathrm{x}+N_\mathrm{y})p_\mathrm{a}/2$	$(N_\mathrm{x}+N_\mathrm{y}+2)P/2$

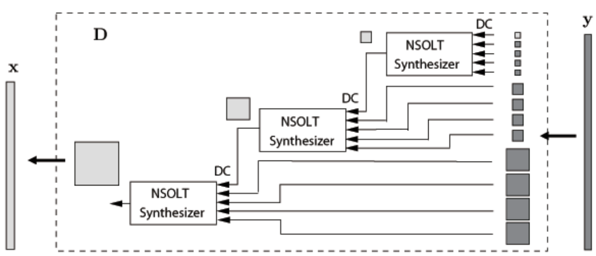


Fig. 2. A tree structure of multiscale NSOLT

constraint can structurally be imposed for NSOLTs in (3) or (4). The condition is met by forcing $\mathbf{W}_0$ to have the form $\mathbf{W}_0 = \left(\begin{smallmatrix} 1 & \mathbf{o} \\ \mathbf{o} & \overline{\mathbf{W}} \end{smallmatrix}\right)$ for the Type-I case and $\mathbf{W}_0 = (\prod_{\ell_\mathrm{y}=1}^{N_\mathrm{y}/2} \mathbf{W}_{\ell_\mathrm{y}}^{\{\mathrm{y}\}} \prod_{\ell_\mathrm{x}=1}^{N_\mathrm{x}/2} \mathbf{W}_{\ell_\mathrm{x}}^{\{\mathrm{x}\}})^{-1} \left(\begin{smallmatrix} 1 & \mathbf{o} \\ \mathbf{o} & \overline{\mathbf{W}} \end{smallmatrix}\right)$ for the Type-II case, where $\overline{\mathbf{W}} \in \mathbb{R}^{(p_\mathrm{s}-1)\times(p_\mathrm{s}-1)}$ is an arbitrary nonsingular matrix [21]. Under the paraunitary constraint, $\overline{\mathbf{W}}$ is chosen to be orthonormal, and $(p_\mathrm{s}-1)$ angular and one sign parameters are reduced from the counts in Table 1.

By iteratively reconstructing the DC components of NSOLT as shown in Fig. 2, a redundant multiscale representation with a Parseval tight frame can be realized. In general, the redundancy $\mathcal{R}$ of dictionary $\mathbf{D} \in \mathbb{R}^{N\times L}$ is determined by the ratio of the numbers of coefficients and pixels as $\mathcal{R} = L/N$. Let τ be the number of tree levels in the multiscale construction and $\mathcal{R}_M^P(\tau)$ be the redundancy of a τ-level tree structure of P-channel NSOLT of downsampling ratio M. Then, the relation

$$\mathcal{R}_M^P(\tau) = \begin{cases} (P-1)\tau + 1, & M = 1 \\ \frac{P-1}{M-1} - \frac{P-M}{(M-1)M^\tau}, & M \geq 2 \end{cases},$$

is verified to hold. The redundancy is guaranteed to be less than $\frac{P-1}{M-1}$ for $M \geq 2$, while it increases proportional to τ for $M = 1$.

3. DICTIONARY LEARNING WITH NSOLT

In this section, we propose a dictionary learning method with a multiscale NSOLT (MS-NSOLT). Let $\mathbf{x}$ be a training image. Note that the proposed dictionaries are trained by using a large region of a given image instead of multiple small subimage patches.

3.1. Problem formulation

We adopt the same goal of MOD and K-SVD to find a dictionary $\mathbf{D}$ and a sparse coefficient vector $\mathbf{y}$ which minimize the approximation error for a given image $\mathbf{x}$ under the K-sparse constraint [12]. This problem is formulated as follows:

$$\{\hat{\mathbf{D}}, \hat{\mathbf{y}}\} \;= \arg\min_{\mathbf{D},\mathbf{y}} \|\mathbf{x} - \mathbf{D}\mathbf{y}\|_2^2 \text{ s.t. } \|\mathbf{y}\|_0 \leq K, \qquad (6)$$

where the expression is simplified for a single image $\mathbf{x}$. The optimization problem in (6) is combinatorial and non-convex. Thus, in the same way to MOD and K-SVD, we divide the learning process into the 'sparse approximation' and 'dictionary update' stage, and alternatively apply these two stages to find a local minimum.

3.2. Sparse approximation stage

The goal of the sparse approximation stage is to find a sparse coefficient vector $\mathbf{y}$ which minimizes the approximation error for a given image $\mathbf{x}$ and a fixed dictionary $\hat{\mathbf{D}}$ under the K-sparse constraint. This problem is formulated as follows:

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

Since this problem is the same as in (2) and NP-hard, we need to adopt a suboptimal approach. Because of the overlapping property of NSOLT, it is preferable to use a sparse coding technique available for large data. IHT is a good candidate for this demand.

(a) *goldhill* (b) *lena* (c) *barbara* (d) *baboon*

Fig. 3. Training images of size 128×128, 8-bit grayscale.

Table 2. Experimental settings of multiscale NSOLT learning

Downsampl. M	2×2
$\sharp$Channels P	8
Polyphase Ord. $\bar{\mathbf{n}}$	$[4, 4]^T$
Sparsity K/N	$2048/128^2$
Initial dictionary	Haar DWT
Training signals	each picture
Sparse Approx.	IHT
Iterations	15

Table 3. Experimental settings of Sparse K-SVD [20]

Block size	8×8
Dictionary size	64×169
Atom sparsity	6
Sparsity K/N	$8/8^2$
Initial dictionary	ODCT
Training signals	10,000
Sparse Approx.	Sparse OMP
Iterations	15

Table 4. PSNRs of the sparse approximation results with learned dictionaries.

Dictionary		*goldhill*	*lena*	*barbara*	*baboon*
Sparse K-SVD ($\mathcal{R} = \frac{169}{64}$)		**33.77**	35.47	35.85	**26.21**
MS-NSOLT ($\mathcal{R} < \frac{8-1}{4-1}$)					
$p_\mathrm{s} + p_\mathrm{a}$	τ				
$4+4$	1	25.80	23.66	22.19	21.44
	2	31.63	35.63	36.59	24.87
	3	32.04	**35.86**	**38.07**	25.64
	4	31.83	35.82	37.64	25.36
	5	31.76	35.53	37.24	25.38
$5+3$	1	24.31	23.01	21.26	21.05
	2	31.74	35.10	36.28	24.89
	3	31.96	34.24	37.22	25.65
	4	31.50	35.52	37.06	25.54
	5	31.85	33.37	36.75	25.29
$6+2$	1	22.83	20.70	20.86	20.38
	2	31.66	35.30	35.90	24.79
	3	31.71	35.39	36.98	25.47
	4	31.90	35.13	36.88	25.36
	5	31.46	34.83	36.33	25.19

3.3. Dictionary update stage

The goal of the dictionary update stage is to find a dictionary $\mathbf{D}$ which minimizes the approximation error for a given image $\mathbf{x}$ and a fixed support of $\tilde{\mathbf{y}}$. This problem is formulated as follows:

$$\{\hat{\mathbf{D}}, \hat{\mathbf{y}}\} = \arg\min_{\mathbf{D},\mathbf{y}} \|\mathbf{x} - \mathbf{D}\tilde{\mathbf{\Omega}}\mathbf{y}\|_2^2, \tag{7}$$

where $\tilde{\mathbf{\Omega}}$ is the diagonal matrix of size $L \times L$ whose diagonal elements represent the positions of nonzero coefficients in $\tilde{\mathbf{y}}$ and defined by $[\tilde{\mathbf{\Omega}}]_{k,k} = 1$ for $[\tilde{\mathbf{y}}]_k \neq 0$ and $[\tilde{\mathbf{\Omega}}]_{k,k} = 0$ for $[\tilde{\mathbf{y}}]_k = 0$. In the same way to SimCO [19], we consider simultaneously updating the dictionary $\mathbf{D}$ and coefficient vector $\mathbf{y}$ by using the relation $\mathbf{y}^* = \mathbf{D}^T\mathbf{x} = \arg\min_{\mathbf{y}} \|\mathbf{x} - \mathbf{D}\mathbf{y}\|_2^2$. Then, the problem in (7) is modified as $\hat{\mathbf{D}} = \arg\min_{\mathbf{D}} \|\mathbf{x} - \mathbf{D}\hat{\mathbf{\Omega}}\mathbf{D}^T\mathbf{x}\|_2^2$.

Since a paraunitary NSOLT can be controlled by angle and sign parameters as in Table 1, the problem is further rewritten as

$$\hat{\mathbf{\Theta}} = \arg\min_{\mathbf{\Theta}} \|\mathbf{x} - \mathbf{D}_{\mathbf{\Theta}}\hat{\mathbf{\Omega}}\mathbf{D}_{\mathbf{\Theta}}^T\mathbf{x}\|_2^2$$

in terms of the parameter set $\mathbf{\Theta} = \{\{\theta_i\}, \{s_i\}\}$, where $\mathbf{D}_{\mathbf{\Theta}}$ is an MS-NSOLT dictionary determined by $\mathbf{\Theta}$. Note that the problem becomes non-linear, although this structured approach keeps the number of design parameters independent of the image size and the number of tree levels, and guarantees the Parseval tight frame property.

4. PERFORMANCE EVALUATION

In order to verify the significance of the proposed method, let us show an example of learned dictionary and evaluate the sparse approximation performances under different settings.

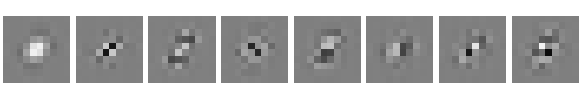


Fig. 4. Learned atoms of size $M_\mathrm{y}(N_\mathrm{y}+1) \times M_\mathrm{x}(N_\mathrm{x}+1) = 10 \times 10$ for *lena* with $\mathbf{M} = \mathrm{diag}(M_\mathrm{y}, M_\mathrm{x}) = \mathrm{diag}(2, 2)$, $M = |\det(\mathbf{M})| = M_\mathrm{y} \times M_\mathrm{x} = 2 \times 2$, $P = p_\mathrm{s} + p_\mathrm{a} = 4+4$, $\bar{\mathbf{n}} = [N_\mathrm{y}, N_\mathrm{x}]^T = [4, 4]^T$ and $\tau = 3$, where single level atoms are shown.

Fig. 3 shows training images in this experiment. We trained dictionaries for each picture independently by changing the number of symmetric channels p_s, that of antisymmetric ones p_a and the tree levels τ. The other settings are summarized in Table 2. The no-DC-leakage property is taken into account and the atom termination at image boundary is adopted [28]. The counts of angle parameters $\sharp\{\theta_i\}$ are 57, 61 and 75 for $p_\mathrm{s} + p_\mathrm{a} = 4+4$, $5+3$ and $6+2$, respectively, and those of sign parameters $\sharp\{s_i\}$ are 39 for every case. In the dictionary update stage, we used the unconstrained minimization function '`fminunc`' of MATLAB R2013b for optimizing angles $\{\theta_i\}$ after initialization of the signs $\{s_i\}$ and angles $\{\theta_i\}$ with the genetic algorithm function '`ga`' of MATLAB R2013b. Fig. 4 shows an example set of learned atoms. As a reference, we will also show the performance of Sparse K-SVD[1]. The settings for learning are summarized in Table 3.

Table 4 shows the PSNRs of the sparse approximation results with the learned dictionaries, where IHT and block-wise Sparse-OMP are used for MS-NSOLT and Sparse K-SVD, respectively. The sparsity is set to remain 12.5% coefficients. From Table 4, it is verified that the proposed method shows comparable or superior performance to the Sparse K-SVD for *lena* and *barbara*. For *goldhill* and *baboon*, however, it shows inferior performance. Our conjecture is that these pictures contain large fine texture regions and may require large fine atoms. To solve this problem, we need further investigation on the dictionary structure, for example the downsampling factor $\mathbf{M}$, number of channels P and polyphase order $\bar{\mathbf{n}}$. In addition, the heavy computational cost due to the non-linear optimization should be reduced since the current approach is not suitable for on-site learning.

5. CONCLUSIONS

A novel design method of 2-D NSOLT was proposed. It was shown that the Parseval tight frame constraint can structurally be imposed during the dictionary learning process. Through the sparse approximation, the significance of the proposal is verified. As future works, we are concerned with the application, further investigation on the structure and efficient implementation of the dictionary update stage.

[1]MATLAB functions in OMPS-Box v1 and KSVDS-Box v11 from `http://www.cs.technion.ac.il/~ronrubin/software.html` were used [20].

6. REFERENCES

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

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

[3] Jean-Luc Starck, Fionn Murtagh, and Jala M. Fadili, *Sparse Image and Signal Processing: Wavelets, Curvelets, Morphological Diversity*, Cambridge University Press, 2010.

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

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

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

[7] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *Proc. of IEEE Asilomar Conf. on Signals, Systems and Computers*, 1993, vol. 1, pp. 40–44.

[8] J. A. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2231–2242, Oct. 2004.

[9] Thomas Blumensath and Mike E. Davies, "Gradient pursuits," *IEEE Trans. Signal Process.*, vol. 56, no. 6, pp. 2370–2382, July 2008.

[10] Thomas Blumensath and Mike E. Davies, "Iterative hard thresholding for compressed sensing," *Appl. Comput. Harmon. Anal.*, vol. 29, pp. 265–274, 2009.

[11] Thomas Blumensath and Mike 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.

[12] Ron Rubinstein, Alfred M. Bruckstein, and Michael Elad, "Dictionaries for sparse representation modeling," *Proc. IEEE*, vol. 98, no. 6, pp. 1045–1057, June 2010.

[13] Michael Unser, "Texture classification and segmentation using wavelet frames," vol. 4, no. 11, pp. 1549, Nov. 1995.

[14] Minh N. Do and Martin Vetterli, "The contourlet transform: An efficient directional multiresolution image representation," *IEEE Trans. Image Process.*, vol. 14, no. 12, pp. 2091–2106, Dec. 2005.

[15] Arthur L. da Cunha, Jianping Zhou, and Minh N. Do, "The nonsubsampled contourlet transform: Theory, design and applications," *IEEE Trans. Image Process.*, vol. 15, no. 10, Oct. 2006.

[16] Ivana Tošić and Pascal Frossard, "Dictionary learning," *IEEE Signal Process. Mag.*, vol. 28, no. 2, pp. 27–38, Mar. 2011.

[17] K. Engan, S. O. Aase, and J. H. Husoy, "Method of optimal directions for frame design," in *Proc. of IEEE ICASSP*, 1999, vol. 5, pp. 2443–2446.

[18] Michal Aharon, Michael Elad, and Alfred Bruckstein, "K-SVD: An algorithm for desigining overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.

[19] Wei Dai, Tao Xu, and Wenwu Wang, "Simultaneous codeword optimization (SimCO) for dictionary update and learning," *IEEE Trans. Signal Process.*, vol. 60, no. 12, pp. 6340–6353, Dec. 2012.

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

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

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

[23] Shogo Muramatsu, Akihiko Yamada, and Hitoshi Kiya, "A design method of multidimensional linear-phase paraunitary filter banks with a lattice structure," *IEEE Trans. Signal Process.*, vol. 47, no. 3, pp. 690–700, Mar. 1999.

[24] Shogo Muramatsu and Dandan Han, "Image denoising with union of directional orthonormal DWTs," in *Proc. of IEEE ICASSP*, Mar. 2012, pp. 1089–1092.

[25] Shogo Muramatsu, Dandan Han, Tomoya Kobayashi, and Hisakazu Kikuchi, "Directional lapped orthogonal transform: Theory and design," *IEEE Trans. Image Process.*, vol. 21, no. 5, pp. 2434–2448, May 2012.

[26] Shogo Muramatsu, Tomoya Kobayashi, Minoru Hiki, and Hisakazu Kikuchi, "Boundary operation of 2-D non-separable linear-phase paraunitary filter banks," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 2314–2318, Apr. 2012.

[27] Natsuki Aizawa, Shogo Muramatsu, and Masahiro Yukawa, "Image restoration with multiple DirLOTs," *IEICE Trans. Fundamentals*, vol. E96-A, no. 10, pp. 1954–1961, Oct. 2013.

[28] Kosuke Furuya, Shintaro Hara, and Shogo Muramatsu, "Boundary operation of 2-D non-separable oversampled lapped transforms," in *Proc. of APSIPA ASC*, Nov. 2013.

[29] P. P. Vaidyanathan, *Multirate Systems and Filter Banks*, Prentice Hall, 1993.

[30] Gilbert Strang and Truong Nguyen, *Wavelets and Filter Banks, Second Edition*, Wellesley Cambridge Pr, 1996.

[31] H. Bolcskei, F. Hlawatsch, and H.G. Feichtinger, "Frame-theoretic analysis of oversampled filter banks," *IEEE Trans. Signal Process.*, vol. 46, no. 12, pp. 3256–3268, Dec. 1998.

[32] Toshihisa Tanaka, "A direct design of oversampled perfect reconstruction FIR filter banks," *IEEE Trans. Signal Process.*, vol. 54, no. 8, pp. 3011–3022, Aug. 2006.

[33] Li Chai, Jingxin Zhang, and Yuxia Sheng, "Optimal design of oversampled synthesis FBs with lattice structure constraints," *IEEE Trans. Signal Process.*, vol. 59, no. 8, pp. 3549–3559, Aug. 2011.