Single-image Super-resolution using Complex Nonseparable Oversampled Lapped Transforms

Satoshi NAGAYAMA*, Shogo MURAMATSU†, Hiroyoshi YAMADA†
* Graduate School of Science and Technology, Niigata Univ., Niigata, Japan
E-mail: satoshi.nagayama.jp@ieee.org
† Faculty of Engineering, Niigata Univ., Niigata, Japan

***Abstract*—In this research, we propose a super-resolution method of complex image using Complex Nonseparable Oversampled Lapped Transforms (ℂNSOLTs). ℂNSOLTs are complex-coefficient transforms based on lattice structure. ℂNSOLTs are possible to design an analysis-synthesis system which simultaneously satisfies nonseparable, redundant, linear-phase, and compact support properties under the constraint of the Parseval tight frame. In addition, ℂNSOLTs are possible to select a redundancy with arbitrary rational number. The effectiveness of the proposed method is evaluated by super-resolution simulation of complex-coefficient images.**

I. Introduction

In many image processing applications, images with high spatial resolution are always required. Image super-resolution (SR) is a method of estimating a high resolution (HR) image from one or more low resolution (LR) images using a signal processing technique [2]–[4]. SR methods are widely used for image processing application where it is difficult to acquire HR images due to limitations of sensing devices or high cost. For instance, synthetic aperture radar (SAR), biomedical imaging, microscopy. SR algorithms can be broadly classified into two classes: single image or multiple image. Many real-world applications require reconstruction of HR image from a single observation. In situations like these, single-image based SR methods may work well.

Many of the recently proposed single-image SRs are based on learning-based algorithms. In the super-resolution based on the learning, training data sets consisting of the LR images and the corresponding HR images are necessary. An HR image can be restored by learning the prior information from the relationship between a training pairs and applying it to the given LR image. Its performance is heavily dependent on the training data set, which requires huge amounts of LR images and HR images for training. Recently, dictionary learning methods based on sparse representation has attracted attention and has achieved high SR performance [1].

Early major study on learning-based SR method using sparse representation has been proposed by Yang *et al.* [5]. They assumed that the local patches extracted from LR images in a training data set can be sparsely represented by consisting redundant dictionary. K-SVD, proposed by Aharon *et al.* [9], is a typical dictionary design method for sparse representation

This work was partially supported by JSPS KAKENHI Grant Numbers JP26420347, JP17H03261.

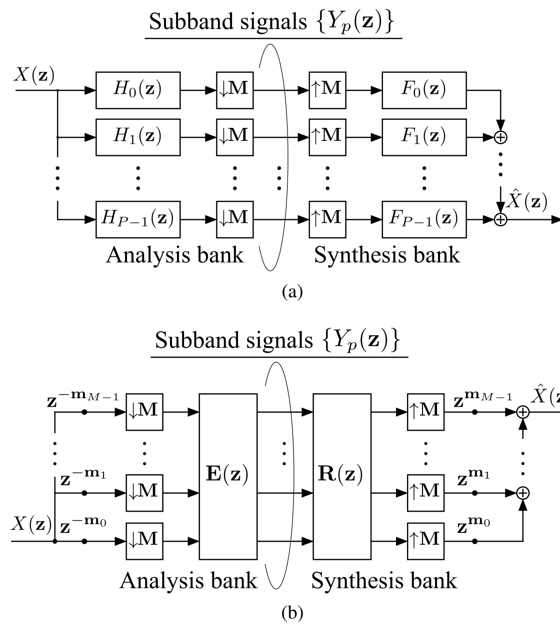


Fig. 1: (a) Parallel Structure of P-chanel filter banks with downsampling factor $\mathbf{M}$ and (b) the polyphase matrix representation, where $\{H_p(\mathbf{z})\}$ and $\{F_p(\mathbf{z})\}$ are analysis and synthesis filters, respectively.

of patches. However, since the K-SVD dictionary does not have useful structural restrictions for image processing such as tight-frame property, invariance, etc, its efficiency is poor as compared with analytical dictionaries.

In [7], we have proposed complex nonseparable oversampled lapped transforms (ℂNSOLTs) for sparse representation of complex-valued images. ℂNSOLTs can construct a redundant analysis/synthesis system with complex-valued atomic images that satisfy non-separable, symmetric, redundant, and compact-support properties using constraints by lattice structure.

Fig. 1 (a) shows a parallel structure of P-channel multidimensional nonseparable filterbanks. The system consists of an analysis and synthesis bank, where $\mathbf{z} \in \mathbb{C}^D$ denotes a $D \times 1$ complex variable vector in the D-dimensional z-transform domain, $H_p(\mathbf{z})$ and $F_p(\mathbf{z})$ are the transfer functions of the p-th analysis and synthesis filter, respectively. In the followings, we consider only the two-dimensional (2-D) separable decimation case, i.e., $D = 2$.

In this work, we propose to use $\mathbb{C}$NSOLT for super-resolution of complex-valued images. In order to show the effectiveness of the proposed method, we compare the performance with the bicubic interpolation method and the undecimated Haar transform.

II. Single-Image Super-Resolution by FISTA

In this section, we review the formulation of SISR problem, and then introduce fast iterative shrinkage/thresholding algorithm as a solver for the problem.

A. Problem Formulation

Fig 2 shows the framework of our problem setting. In this model, $\mathbf{x} \in \mathbb{C}^N$ is the observed LR image. Let $\mathbf{u} \in \mathbb{C}^M$ $(M > N)$ be an unknown original HR image and

$$\mathbf{x} = \mathbf{SHu} + \mathbf{w}, \tag{1}$$

where $\mathbf{H} \in \mathbb{C}^{M\times M}$ is a blur operator, $\mathbf{S} \in \mathbb{R}^{N\times M}$ is a decimation operator, and $\mathbf{w} \in \mathbb{C}^N$ is a additive noise. This SR problem is treated as a problem to estimate the unknown original image $\mathbf{u}$ from the known observation image $\mathbf{x}$. Under these conditions, the estimation method using the sparseness improves the performance. We assume that the original image $\mathbf{u}$ can be expressed as a linear combination of a small number of element images, that is

$$\mathbf{u} = \mathbf{Dy}. \tag{2}$$

where $\mathbf{y} \in \mathbb{C}^L$ is a coefficient vector and $\mathbf{D} \in \mathbb{C}^{M\times L}$ is an dictionary. The super-resolution problem is replaced by the problem of estimating the sparse coefficient vector $\mathbf{y}$ and formulated as ℓ_1-norm regularized least squares problem

$$\hat{\mathbf{y}} = \arg\min_{\mathbf{y}} \|\mathbf{x} - \mathbf{SHDy}\|_2^2 + \lambda\|\mathbf{y}\|_1, \tag{3}$$

where $\lambda \in \mathbb{R}_+$ is the trade-off parameter between the first and second term in the right hand side, i.e., fidelity and sparsity.

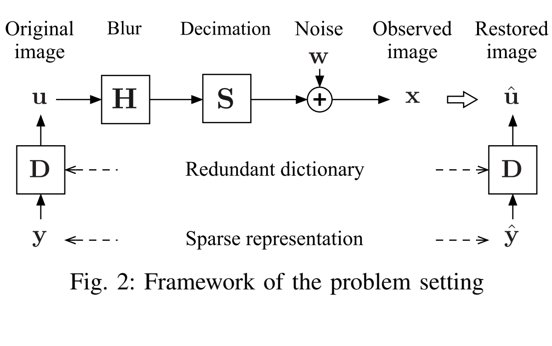


Fig. 2: Framework of the problem setting

Algorithm 1 Image restoration with FISTA

Data: Observed image $\mathbf{x} \in \mathbb{C}^N$
Result: Restored image $\hat{\mathbf{u}} \in \mathbb{C}^M$
{initialization}
$t \leftarrow 0$
$\mathbf{y}^{(0)} \leftarrow \mathbf{A}^H\mathbf{x}$
$\mathbf{v}^{(1)} \leftarrow \mathbf{y}^{(0)}$
$s_1 \leftarrow 1$
Find $\mathbf{y}$ that minimizes $\{\frac{1}{2}\|\mathbf{x} - \mathbf{Ay}\|_2^2 + \|\mathbf{y}\|_1\}$
repeat
  $t \leftarrow t + 1$
  $\mathbf{y}^{(t)} \leftarrow \mathcal{T}_{\frac{\lambda}{L}}\left(\mathbf{v}^{(t)} + \frac{1}{L}\mathbf{A}^H\left(\mathbf{x} - \mathbf{Av}^{(t)}\right)\right)$
  $s_{t+1} \leftarrow (1 + \sqrt{1 + 4s_t^2})/2$
  $\tilde{s}_{t+1} \leftarrow (s_t - 1)/s_{t+1}$
  $\mathbf{v}^{(t+1)} \leftarrow \mathbf{y}^{(t)} + \tilde{s}_{t+1}(\mathbf{y}^{(t)} - \mathbf{y}^{(t-1)})$
until $\|\mathbf{y}^{(t)} - \mathbf{y}^{(t-1)}\|_2^2/\|\mathbf{y}^{(t)}\|_2^2 < \epsilon$
$\hat{\mathbf{u}} \leftarrow \mathbf{Dy}^{(t)}$

B. Fast Iterative shrinkage/thresholding algorithm (FISTA)

If a vector $\mathbf{y} = [y_0, y_1, \cdots, y_{L-1}]^T$ is complex-valued, ℓ_1-norm regularization term in (3) can be rewritten by linear combination of ℓ_2-norms of real-valued vectors as

$$\|\mathbf{y}\|_1 = \sum_{\ell=0}^{L-1} \sqrt{\mathfrak{Re}(y_l)^2 + \mathfrak{Im}(y_\ell)^2} = \sum_{\ell=0}^{L-1} \|\mathbf{g}_\ell\|_2, \tag{4}$$

where $\mathbf{g}_\ell = (\mathfrak{Re}(y_\ell), \mathfrak{Im}(y_\ell))^T$.

Algorithm 1 shows the FISTA for complex-valued signals, where $(\cdot)^H$ denotes the Hermitian transpose of matrix. $\mathcal{T}_\lambda(\mathbf{y})$ is the soft shrinkage operator defined by

$$\mathcal{T}_\lambda(\mathbf{v}) = (\mathbf{1} - \lambda\,(\mathbf{1} \oslash |\mathbf{v}|))_+ \odot \mathbf{v}, \tag{5}$$

where $\odot, \oslash$ are vector operators that represent element-wise multiplication and division, respectively, and operator $|\cdot| : \mathbb{C}^N \to \mathbb{R}_+^N$ takes element-wise absolute value.

III. Review of Complex Nonseparable Oversampled Lapped Transforms

A selection of dictionary $\mathbf{D}$ affects the performance of SR method in (3). In this section, we propose to use $\mathbb{C}$NSOLTs as the dictionary $\mathbf{D}$.

A. Lattice Structure of 2-D Type-I $\mathbb{C}$NSOLT

A lattice structure of 2-D $\mathbb{C}$NSOLT analysis bank $\mathbf{E}(\mathbf{z})$ is given by cascade of matrices $\{\mathbf{G}_k^{\{d\}}(z_d)\}$ as

$$\mathbf{E}(\mathbf{z}) = \mathbf{\Phi}\left(\prod_{k=1}^{K_\mathrm{v}} \mathbf{G}_k^{\{\mathrm{v}\}}(z_\mathrm{v})\right)\left(\prod_{k=1}^{K_\mathrm{h}} \mathbf{G}_k^{\{\mathrm{h}\}}(z_\mathrm{h})\right) \cdot \mathbf{E}_0(\mathbf{z}), \tag{6}$$

where $\mathbf{E}_0(\mathbf{z}) \in \mathbb{C}^{P\times M}[\mathbb{C}^D]$ is a PU initial matrix and $\{\mathbf{G}_k^{\{d\}}(z_d) \in \mathbb{C}^{P\times P}[\mathbb{C}^D]\}$ are propagation matrices of which d-th dimension order is $N_{\mathrm{G},d} \in \mathbb{N}$ and the other is 0, $\mathbf{\Phi} \in \mathbb{C}^{P\times P}$ is a diagonal matrix related to symmetry of filters and its each element is ± 1 or $\pm j$.

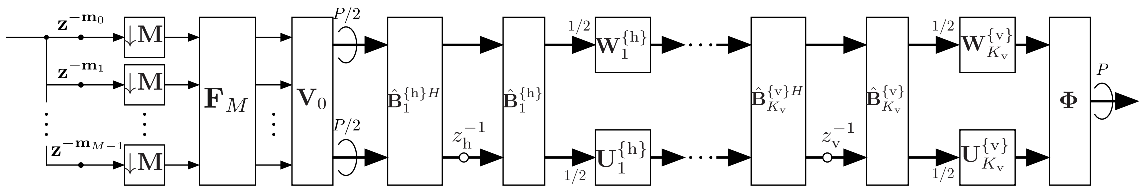


Fig. 3: A lattice structure of P-channel 2-D Type-I $\mathbb{C}$NSOLT with decimation factor $\mathbf{M}$

The propagation matrix of Type-I $\mathbb{C}$NSOLT is given by

$$\mathbf{G}_k^{\{d\}}(z_d) = \mathbf{V}_k^{\{d\}}\mathbf{Q}_k^{\{d\}}(z_d) \tag{7}$$

where

$$\mathbf{V}_k^{\{d\}} = \begin{pmatrix} \mathbf{W}_k^{\{d\}} & \mathbf{O} \\ \mathbf{O} & \mathbf{U}_k^{\{d\}} \end{pmatrix} \tag{8}$$

$$\mathbf{Q}_k^{\{d\}}(z_d) = \hat{\mathbf{B}}_k^{\{d\}}\mathbf{\Lambda}(z_d)\hat{\mathbf{B}}_k^{\{d\}H} \tag{9}$$

$$\mathbf{\Lambda}(z_d) = \begin{pmatrix} \mathbf{I}_{\lfloor \frac{P}{2} \rfloor} & \mathbf{O} \\ \mathbf{O} & z_d^{-1}\mathbf{I}_{\lfloor \frac{P}{2} \rfloor} \end{pmatrix} \tag{10}$$

$$\hat{\mathbf{B}}_k^{\{d\}} = \frac{1}{\sqrt{2}} \begin{pmatrix} \hat{\mathbf{C}}_k^{\{d\}} & \hat{\mathbf{C}}_k^{\{d\}*} \\ \hat{\mathbf{S}}_k^{\{d\}} & \hat{\mathbf{S}}_k^{\{d\}*} \end{pmatrix}. \tag{11}$$

$$\hat{\mathbf{C}}_k^{\{d\}} = \begin{cases} \mathrm{diag}\left(\hat{\mathbf{c}}_{k,0}^{\{d\}}, \hat{\mathbf{c}}_{k,1}^{\{d\}}, \cdots, \hat{\mathbf{c}}_{k,\lfloor \frac{P}{4} \rfloor-1}^{\{d\}}\right), & \lfloor \frac{P}{4} \rfloor \text{ even} \\ \mathrm{diag}\left(\hat{\mathbf{c}}_{k,0}^{\{d\}}, \hat{\mathbf{c}}_{k,1}^{\{d\}}, \cdots, \hat{\mathbf{c}}_{k,\lfloor \frac{P}{4} \rfloor-1}^{\{d\}}, 1\right), & \lfloor \frac{P}{4} \rfloor \text{ odd} \end{cases} \tag{12}$$

$$\hat{\mathbf{S}}_k^{\{d\}} = \begin{cases} \mathrm{diag}\left(\hat{\mathbf{s}}_{k,0}^{\{d\}}, \hat{\mathbf{s}}_{k,1}^{\{d\}}, \cdots, \hat{\mathbf{s}}_{k,\lfloor \frac{P}{4} \rfloor-1}^{\{d\}}\right), & \lfloor \frac{P}{4} \rfloor \text{ even} \\ \mathrm{diag}\left(\hat{\mathbf{s}}_{k,0}^{\{d\}}, \hat{\mathbf{s}}_{k,1}^{\{d\}}, \cdots, \hat{\mathbf{s}}_{k,\lfloor \frac{P}{4} \rfloor-1}^{\{d\}}, j\right), & \lfloor \frac{P}{4} \rfloor \text{ odd} \end{cases} \tag{13}$$

$$\hat{\mathbf{c}}_{k,p}^{\{d\}} = \begin{pmatrix} -j\cos\theta_{k,p}^{\{d\}} & -j\sin\theta_{k,p}^{\{d\}} \\ \cos\theta_{k,p}^{\{d\}} & -\sin\theta_{k,p}^{\{d\}} \end{pmatrix} \tag{14}$$

$$\hat{\mathbf{s}}_{k,p}^{\{d\}} = \begin{pmatrix} \sin\theta_{k,p}^{\{d\}} & \cos\theta_{k,p}^{\{d\}} \\ j\sin\theta_{k,p}^{\{d\}} & -j\cos\theta_{k,p}^{\{d\}} \end{pmatrix}, \tag{15}$$

where the superscript "$*$" denotes complex-conjugate, $\mathbf{W}_k^{\{d\}}, \mathbf{U}_k^{\{d\}} \in \mathbb{R}^{\lfloor \frac{P}{2} \rfloor \times \lfloor \frac{P}{2} \rfloor}$ are invertible matrix, and $\theta_{k,p}^{\{d\}} \in \mathbb{R}$ is the angle parameter.

The initial matrix is given by

$$\mathbf{E}_0(\mathbf{z}) = \mathbf{V}_0\mathbf{F}_\mathbf{M}\mathbf{J}_M. \tag{16}$$

Where $\mathbf{F}_\mathbf{M} \in \mathbb{C}^{M \times M}$ is any $M \times M$ unitary matrix that satisfies $\mathbf{F}_\mathbf{M}^*\mathbf{J} = \mathbf{F}_\mathbf{M}$. $\mathbf{V}_0 \in \mathbb{R}^{P \times M}$ is an arbitrary left-invertible matrix. We adopt a centered discrete Fourier transform (CDFT) matrix as $\mathbf{F}_\mathbf{M}$ [10]. The p-th row and m-th column element of 1-D M-point CDFT matrix is defined as

$$f_{m,p} = \frac{1}{\sqrt{M}}\exp\left(-j\pi\frac{p(2m+1)}{M}\right) \tag{17}$$

in the element-wise representation.

B. Dictionary Learning

The dictionary $\mathbf{D}$ based on $\mathbb{C}$NSOLT can be realized with synthesis filter bank $\mathbf{R}(\mathbf{z}) = \mathbf{z}^{-\bar{\mathbf{n}}}\tilde{\mathbf{E}}(\mathbf{z})$ and can be designed by example-based learning. Here, $\tilde{\mathbf{E}}(\mathbf{z})$ is paraconjugate of analysis bank $\mathbf{E}(\mathbf{z})$ and $\bar{\mathbf{n}}$ is the polyphse order vector [8]. The dictionary learning method is formulated as

$$\left\{\hat{\mathbf{D}}, \{\hat{\mathbf{y}}_i\}\right\} = \arg\min_{\mathbf{D},\{\mathbf{y}_i\}} \frac{1}{S}\sum_{i=0}^{S-1}\|\mathbf{x}_i - \mathbf{D}\mathbf{y}_i\|_2^2 \text{ s.t. } \|\mathbf{y}_i\|_0 \leq K, \tag{18}$$

where $\|\cdot\|_0$ denotes ℓ_0-norm, i.e., the number of non-zero elements, and $\{\mathbf{x}_i\}$ denotes the training images. A typical dictionary learning method consists of two steps: "sparse approximation" and "dictionary update" stage. These two processes are repeatedly performed in order to obtain an optimum dictionary.

- *Sparse approximation stage* finds coefficient vector $\mathbf{y}_i$ such that the approximation error is minimized on the training image $\mathbf{x}_i$ under the K-sparse constraint. This problem is formulated as

$$\hat{\mathbf{y}}_i = \arg\min_{\mathbf{y}_i}\|\mathbf{x}_i - \mathbf{D}\mathbf{y}_i\|_2^2 \text{ s.t. } \|\mathbf{y}_i\|_0 \leq K. \tag{19}$$

- *Dictionary learning stage* finds dictionary $\mathbf{D}$ such that the approximation error is minimized on the training image $\mathbf{x}_i$ $\mathbb{C}$NSOLT dictionary can be designed by controlling angle parameter vectors $\boldsymbol{\phi} = (\phi_0, \phi_1, \cdots, \phi_{P-1})^T$ and $\boldsymbol{\theta} = (\theta_0, \theta_1, \cdots, \theta_{J-1})^T$, and sign parameters $\mathbf{s} = (s_0, s_1, \cdots, s_{I-1})^T$, where $\boldsymbol{\theta}, \mathbf{s}$ are obtained by Givens rotations factorization form of parameter matrices of $\mathbb{C}$NSOLT. This optimization problem is formulated as

$$\hat{\mathbf{\Theta}} = \arg\min_{\mathbf{\Theta}} \frac{1}{S}\sum_{i=0}^{S-1}\|\mathbf{x}_i - \mathbf{D}\mathbf{y}_i\|_2^2, \tag{20}$$

where $\mathbf{D}_\mathbf{\Theta}$ is the $\mathbb{C}$NSOLT dictionary given by $\mathbf{\Theta} = \{\boldsymbol{\phi}, \boldsymbol{\theta}, \mathbf{s}\}$. The updated dicionary $\mathbf{D} = \mathbf{D}_\mathbf{\Theta}$ is used in the sparse approximation stage in the next iteration.

IV. Performance Evaluation

In order to show the effectiveness of the proposed method, super-resolution simulation of a complex image is conducted and comparison with existing method is performed. PSNR is used as an evaluation index of the super-resolution image. Bicubic interpolation (BCI) and undecimated Haar transformation (UDHT) are used as comparison objects for performance evaluation. In the observation process, we set the operator $\mathbf{H}$

as Gaussian blur with $\sigma^2_{\rm blur} = 2.00 \times 10^{-4}$, and the operator $\mathbf{S}$ as 2×2 decimation matrix. We use the FISTA as the solver for the super-resolution problem (3) except for BCI. Table I shows the configurations of ℂNSOLT and UDHT. Fig. 5 shows the original LR image and observed HR image for the simulation. Fig. 4 shows the atomic images of the ℂNSOLT dictionary.

Simulation results are shown in Fig. 6. Compared to BCI, the proposed method improves PSNR by about 8[dB]. In addition, the proposed method achieves low redundancy while realizing the same degree of restoration performance as UDHT.

V. Conclusion

In this research, we proposed a super-resolution method of complex image using ℂNSOLT. By comparing super-resolution performance with BCI and FISTA with UDHT, the effectiveness of the proposed method was shown. We will evaluate the performance against actual millimeter wave radar images in the future.

References

[1] M. Elad, *Sparse and Redundant Representations.* New York: Springer, 2010.

[2] K. Nasrollahi and T. Moeslund, "Super-resolution: a comprehensive survey," *Machine Vision and Applications,* vol. 25, no. 6, pp. 1423–1468, 2014.

[3] C. Jiji, S. Chaudhuri, and P. Chatterjee "Single frame image super-resolution: should we process locally or globally?," *Multidimensional Systems and Signal Proessing,* vol. 18, no. 2–3, pp. 123–152, 2007.

[4] E. Candés and C. Fernandez-Granda, "Super-resolution from noisy data," *Journal of Fourier Analysis and Applications,* vol. 19, no. 6, pp. 1229-1254, 2013.

[5] J Yang, J. Wright, T. S. Huang, and Y. Ma, "Image super-resolution via sparse representation," vol. 19, no. 11, pp. 2861–2873, 2010.

[6] S. Nagayama, S. Muramatsu, H. Yamada, and Y. Sugiyama, "Millimeter wave radar image denoising with complex nonseparable oversampled lapped transform," *Proc. of APSIPA ASC,*, December 2017.

[7] S. Nagayama, S. Muramatsu, H. Yamada, and Y. Sugiyama, "Complex nonseparable oversampled lapped transform for sparse representation of millimeter wave radar image," *Proc. of International Conference on Image Processing (ICIP),* pp. 2716–2720, September 2017.

[8] S. Muramatsu, K. Furuya, and N. Yuki "Multidimensional nonseparable oversampled lapped transforms: theory and design," *IEEE Trans. on Signal Processing,* vol. 65, no. 5, pp. 1251–1264, March 2017.

TABLE I: simulation configuration of the dictionaries

	ℂNSOLT	UDHT
# channels P	6	4
Decimation factor $(M_{\rm h}, M_{\rm v})$	$(2,2)$	$(1,1)$
Polyphase order$(N_{\rm h}, N_{\rm v})$	$(2,2)$	$(1,1)$
Tree Levels τ	2	2
Redundancy R	≈ 1.63	7

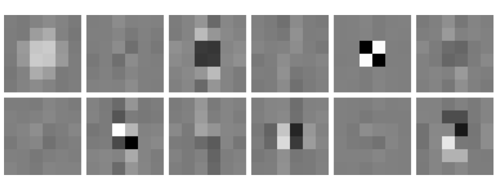
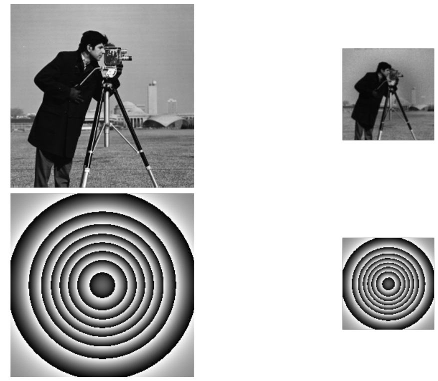
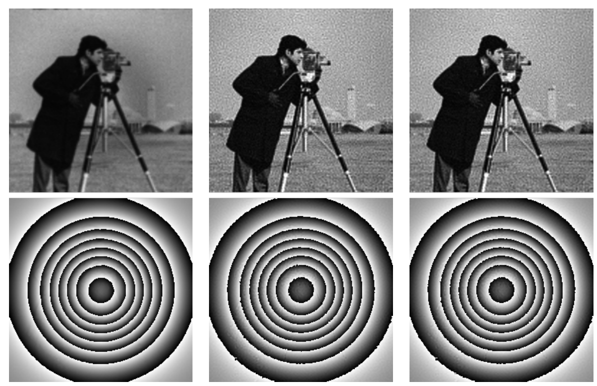


Fig. 4: Atomic images of ℂNSOLT, the upper part shows the real part and lower part shows the imaginary part



(a) Original Image 256×256[pixel] (b) Observed Image 128×128[pixel]

Fig. 5: (a) Original Image and (b) Observed Image. The upper image shows square of the magnitude, the lower one shows the argument



(a) (b) (c)

Fig. 6: Results of super-resolution simulation (a) BCI, PSNR = 17.521[dB], $\lambda = 1 \times 10^{-4}$. (b) UDHT, PSNR = 25.334[dB], $\lambda = 1 \times 10^{-4}$. (c) ℂNSOLT, PSNR = 25.435[dB]

[9] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: an algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. on Signal Processing,* vol. 54, no. 11, pp. 4311–4322, 2006.

[10] D. Mugler, "The centered discrete fourier transform and a parallel implementation of the FFT," *Proc. of Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP),* pp. 1725–1728, May 2011.