RESTORATION OF IMAGES AND 3D DATA TO HIGHER RESOLUTION BY DECONVOLUTION WITH SPARSITY REGULARIZATION

Yingsong Zhang, Nick Kingsbury

Signal Processing & Communication Laboratory
Department of Engineering, University of Cambridge, U.K.
{yz304, ngk10}@cam.ac.uk

ABSTRACT

Image convolution is conventionally approximated by the LTI discrete model. It is well recognized that the higher the sampling rate, the better is the approximation. However sometimes images or 3D data are only available at a lower sampling rate due to physical constraints of the imaging system. In this paper, we model the under-sampled observation as the result of combining convolution and subsampling. Because the wavelet coefficients of piecewise smooth images tend to be sparse and well modelled by tree-like structures, we propose the L0 reweighted-L2 minimization ($\mathrm{L_0RL_2}$) algorithm to solve this problem. This promotes sparsity by minimizing the reweighted L2 norm, which approximates the L0 norm, and by enforcing a tree model through bivariate shrinkage. We test the algorithm on 3 examples: a simple ring, the cameraman image and a 3D microscope dataset; and show that good results can be obtained.

Index Terms— Image restoration, deconvolution, sparsity, L0 norms, regularization.

1. INTRODUCTION

Usually convolution is assumed to be a continuous linear shift-invariant system $y = h \otimes x + n$, which is often approximated by the discrete vector model

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}.$$

It is well recognized that higher sampling rates lead to better approximations. However, sometimes only a relatively low sampling rate can be achieved due to physical constraints in the system hardware or to measurement time constraints.

In some situations the point-spread function (psf) $\mathbf{H}$ at a higher sampling rate is available, although $\mathbf{y}$ at the same sampling rate is not. For example the psf $\mathbf{H}$ of a 3D microscope can be calculated or well calibrated although the image data must be sampled at a lower rate due to physical / temporal constraints. (Often the under-sampling is most pronounced in the inter-slice or z direction.) Hence, a question arises: can we recover the ideal sampled $\mathbf{x}$ from the under-sampled $\mathbf{y}$, using the known psf at a higher sampling rate?

Assume $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}$ is the discrete model obtained by ideal sampling, and the observation is only available at a lower sampling rate, which we model as:

$$\mathbf{D}(\mathbf{y}) = \mathbf{D}(\mathbf{H}\mathbf{x}) + \mathbf{D}(\mathbf{n})$$

where $\mathbf{D}$ is a matrix that represents the subsampling operation. For simplicity, we denote the undersampled observations and noise as $\bar{\mathbf{y}} = \mathbf{D}(\mathbf{y})$ and $\bar{\mathbf{n}} = \mathbf{D}(\mathbf{n})$ to give:

$$\bar{\mathbf{y}} = \mathbf{D}(\mathbf{H}\mathbf{x}) + \bar{\mathbf{n}} \tag{1}$$

Then the problem can be expressed as recovery of the higher-sampled $\mathbf{x}$ from $\bar{\mathbf{y}}$, with known PSF $\mathbf{H}$ and subsampling operation $\mathbf{D}$ according to eq(1).

2. SPARSE PENALTY IN THE WAVELET DOMAIN

Wavelet decompositions provide insight about image structure and scale. They decompose signals or images into different scales, where the scales produce a pyramid structure. Moreover, the coefficients show strong persistence across scales; i.e. large parent coefficients tend to have large child coefficients while small parents tend to have small children. This connection between parents and children is well known as the tree structure of wavelet transforms, and it has been much exploited for image modeling and denoising [2, 3, 4]. The inter-scale correlations are particularly strong for the magnitudes of coefficients from analytic complex wavelets from transforms such as the dual tree (DT $\mathbb{C}$WT) [1].

Besides the tree-structure property, the wavelet representation of piecewise smooth images is well known to be compressible [5], as exploited in the now well-established sparsity regularization approaches developed in the context of compressive sensing [6]. Recently, Baraniuk et al. proposed model-based compressive sensing, which utilizes sparsity and tree-structure simultaneously to improve the recovered signal quality [7].

3. MODEL-BASED RECOVERY ALGORITHM

In this paper we modify the DT $\mathbb{C}$WT-based L0 reweighted-L2 minimization algorithm ($\mathrm{L_0RL_2}$ [1])[8] to solve the problem, because the algorithm is simple enough to be used with

[1] the $\mathrm{L_0RL_2}$ algorithm was previously called ℓ_{02} in [9]

large 3D datasets and it is able to recover much of the missing information in fine-scale subbands [9].

Firstly, we use real matrices $\mathbf{W}$ and $\mathbf{M}$ to represent the forward and inverse DT $\mathbb{C}$WT, and a column vector $\mathbf{w}$ to denote the wavelet coefficients of the transform of the image $\mathbf{x}$, such that $\mathbf{w} = \mathbf{W}\mathbf{x}$, and $\mathbf{x} = \mathbf{M}\mathbf{w}$ is the image reconstructed from $\mathbf{w}$. w_i is the ith entry of $\mathbf{w}$. We denote by $(\mathbf{W}_j)_{j\in S}$ and $(\mathbf{M}_j)_{j\in S}$ the different decomposition and reconstruction wavelet subspaces (subbands) respectively, where $S = 0, 1, \dots J$. Note that J is the number of wavelet subspaces, and 0 indexes the scaling function subspace. Specifically $\mathbf{W}_j$ and $\mathbf{M}_j$ are masked versions of $\mathbf{W}$ and $\mathbf{M}$, respectively, such that $\mathbf{w} = \sum_{j\in S}\mathbf{w}_j = \sum_{j\in S}\mathbf{W}_j\mathbf{x}$ and $\mathbf{M} = \sum_{j\in S}\mathbf{M}_j$. For simplicity, we denote $\mathbf{W}_j\mathbf{x}$ as column vector $\mathbf{w}_j$, with non-zero coefficients only in subband j.

3.1. L0 reweighted-L2 minimization

To measure sparseness, we use the weighted L2-norm to approximate the L0 sparseness measure:

$$\sum_i a_i|w_i|^2 = \sum_i \frac{1}{|w_i|^2+\epsilon^2}|w_i|^2 \approx \|\mathbf{w}\|_0 \qquad (2)$$

where ϵ serves as a stabilizer to avoid infinity and set the threshold between wavelet coefficients that are regarded as '0' and those regarded as '1'. Typically, ϵ^2 is of the order of σ^2, the variance of the measurement noise $\mathbf{n}$.

As in [8], we formulate the primary penalty function as:

$$J(\mathbf{w}) = \|\bar{\mathbf{y}} - \mathbf{D}(\mathbf{H}\mathbf{M}\mathbf{w})\|^2 + \sigma^2\lambda\mathbf{w}^T\mathbf{A}\mathbf{w} \qquad (3)$$

where the weight matrix $\mathbf{A}$ is diagonal with $A_{ii} = a_i$, and λ controls the level of sparsity.

Then, employing the subband pre-emphasis process from [8, 11], we have the following auxiliary penalty function:

$$\begin{aligned}\bar{J}_n(\mathbf{w}) \;=\; & J(\mathbf{w}) + \sum_{j\in S}\alpha_j\|\mathbf{W}_j\mathbf{x}^{(n)} - \mathbf{w}_j\|^2 \\ & - \|\mathbf{D}\mathbf{H}\mathbf{x}^{(n)} - \mathbf{D}\mathbf{H}\mathbf{M}\mathbf{w}\|^2 \end{aligned} \qquad (4)$$

where $\mathbf{x} = \mathbf{M}\mathbf{w} = \sum_{j\in S}\mathbf{M}_j\mathbf{w}_j$, $\mathbf{x}^{(n)}$ is the n^{th} iteration of $\mathbf{x}$ and the α_j must be properly chosen to make sure that $\bar{J}_n(\mathbf{w}) - J(\mathbf{w}) \geq 0$ (see [8] for "the setting of $\boldsymbol{\alpha}$").

Let $\Phi = \mathbf{D}\mathbf{H}$, and let $C(\mathbf{x}^{(n)}, \bar{\mathbf{y}}) = \|\bar{\mathbf{y}}\|^2 - \|\Phi\mathbf{x}^{(n)}\|^2$ which is a function that does not depend on $\mathbf{w}_j$. We can then rewrite eq(4) as a quadratic equation in $\mathbf{w}_j$:

$$\begin{aligned}&\bar{J}_n(\mathbf{w}_1, \dots, \mathbf{w}_J) \;= \\ &-2(\Phi^T\bar{\mathbf{y}})^T\sum_{j\in S}\mathbf{M}_j\mathbf{w}_j + \sum_{j\in S}\alpha_j\|\mathbf{W}_j\mathbf{x}^{(n)} - \mathbf{w}_j\|^2 \\ &+2(\Phi^T\Phi\mathbf{x}^{(n)})^T\sum_{j\in S}\mathbf{M}_j\mathbf{w}_j + \sigma^2\lambda\sum_{j\in S}\mathbf{w}_j^T\mathbf{A}_j\mathbf{w}_j + C(\mathbf{x}^{(n)}, \bar{\mathbf{y}})\end{aligned} \qquad (5)$$

whose minimum is simply obtained by setting

$$\partial\bar{J}_n(\mathbf{w}_1, \dots, \mathbf{w}_J)/\partial\mathbf{w}_j = 0,$$

and the solution (using $\mathbf{M} = \mathbf{W}^T$) is, for any $j \in S$:

$$(\alpha_j + \sigma^2\lambda\mathbf{A}_j)\mathbf{w}_j = \mathbf{W}_j\left(\alpha_j\mathbf{x}^{(n)} + \Phi^T(\bar{\mathbf{y}} - \Phi\mathbf{x}^{(n)})\right) \qquad (6)$$

where $\mathbf{A}_j$ is a masked version of $\mathbf{A}$ with non-zero coefficients only in subband j.

Hence we get the following algorithm ($\mathrm{L_0RL_2}$):

INPUT: Measurement $\bar{\mathbf{y}}$, noise variance σ^2, initial $\mathbf{x}^{(0)}$.
INITIALIZE: Set the weights:
$\mathbf{w} = \mathbf{W}\mathbf{x}^{(0)},\ A_{ii} = \left(|w_i|^2 + \epsilon^2\right)^{-1}$
Repeat the following until convergence or a fixed number of times
APPROX: $\mathbf{z}_j = \mathbf{W}_j\left(\alpha_j\mathbf{x}_j^{(n)} + \mathbf{H}^T\mathbf{D}^T(\bar{\mathbf{y}} - \mathbf{D}\mathbf{H}\mathbf{x}^{(n)})\right)$
and threshold according to the structure set by $\mathbf{A}$
$\mathbf{x}^{(n+1)} = \sum_j \mathbf{M}\left((\alpha_j + \sigma^2\lambda\mathbf{A}_j)^{-1}\mathbf{z}_j\right)$.
UPDATE: After several approximations, update the weights using estimate $\mathbf{x}^{(n+1)}$:
$\mathbf{w} = \mathbf{W}\mathbf{x}^{(n+1)},\ A_{ii} = \left(|w_i|^2 + \epsilon^2\right)^{-1}$

The initial guess $\mathbf{x}^{(0)}$ can be obtained by $\mathbf{D}^{-1}(\mathbf{D}\mathbf{H})^{-1}\bar{\mathbf{y}}$.

Because the missing data in observations can be treated similarly to the effect of noise, λ should be set larger than 1 to encourage the regeneration of such missing information. As the iterations proceed and missing data is progressively filled in, λ is slowly decreased to 1 so that $\sigma^2\lambda$ reaches the actual noise level.

3.2. Embedding tree models into the algorithm

The above $\mathrm{L_0RL_2}$ algorithm uses the weight matrix $\mathbf{A}$ as a regularization constraint to confine the support of the large coefficients. We can expect good results if $\mathbf{A}$ is accurate enough. The tree model in the wavelet domain has been intensively researched, and many tree-model based algorithms have been proposed and successfully applied both for denoising [2, 3, 4] and compressive sensing [7, 10].

These methods can be used in our algorithm to find a proper $|w_i|^2$. Many of these methods can be computationally demanding, so in the following experiments we use bivariate shrinkage [4] to denoise $\mathbf{W}\mathbf{x}^{(n+1)}$ and then use the denoised coefficients to update $\mathbf{A}$. We find that bivariate shrinkage gives a good tradeoff between performance and computational efficiency.

4. EXPERIMENTAL RESULTS

The first experiment (Fig.1) aims to show that better interpolation can be obtained by the nonlinear $\mathrm{L_0RL_2}$ algorithm compared to linear spline interpolation. The original image is a ring with an inner radius of 80 pixels and a width of 6 pixels with a small amount of edge smoothing to reduce aliasing effects. The whole image is of size 256×256. Only the odd columns of the image are preserved and the even columns are discarded, which simulates in 2D the effect of missing slices in 3D data. Then the full-resolution image is reconstructed, first using spline interpolation and then using the proposed $\mathrm{L_0RL_2}$ algorithm. Only ten iterations of $\mathrm{L_0RL_2}$ are needed to

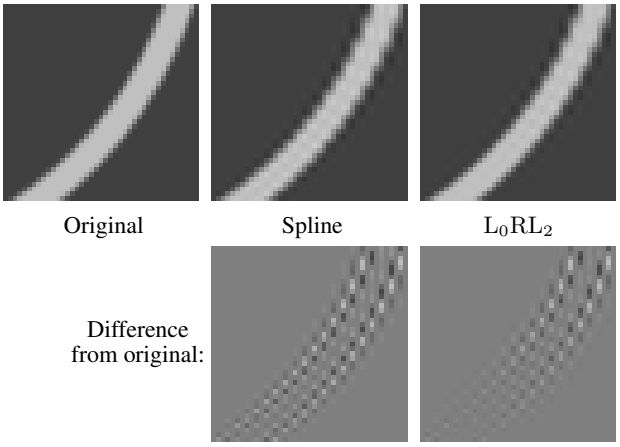


Fig. 1. Reconstruction of part of a ring of unit intensity change, when alternate columns are discarded. In the difference images, midgray represents 0, black -0.5, white 0.5.

achieve the result shown. Note how the L_0RL_2 error becomes very small when the angle of the ring is below 45 degrees.

In figure 2, we show a deconvolution experiment on the standard 2D test image Cameraman, which has been subject to blur and added noise, and alternate columns have been discarded. To demonstrate the effectiveness of the new model (eq(1)), we compare it with deconvolving the subsampled and blurred image first, followed by spline interpolation to reconstruct the missing columns.

For figure 2, we convolved the image with a 16×16 Gaussian blurring kernel with $\sigma = 1$ pel, and added noise at BSNR = 40 dB. We used the under-regularized Wiener filter to obtain the initial estimate:

$$\mathbf{D}(\mathbf{x}^{(0)}) = (\mathbf{H}^T\mathbf{D}^T\mathbf{D}\mathbf{H} + 10^{-3}\sigma^2\mathbf{I})^{-1}(\mathbf{D}\mathbf{H})^T\bar{\mathbf{y}}$$

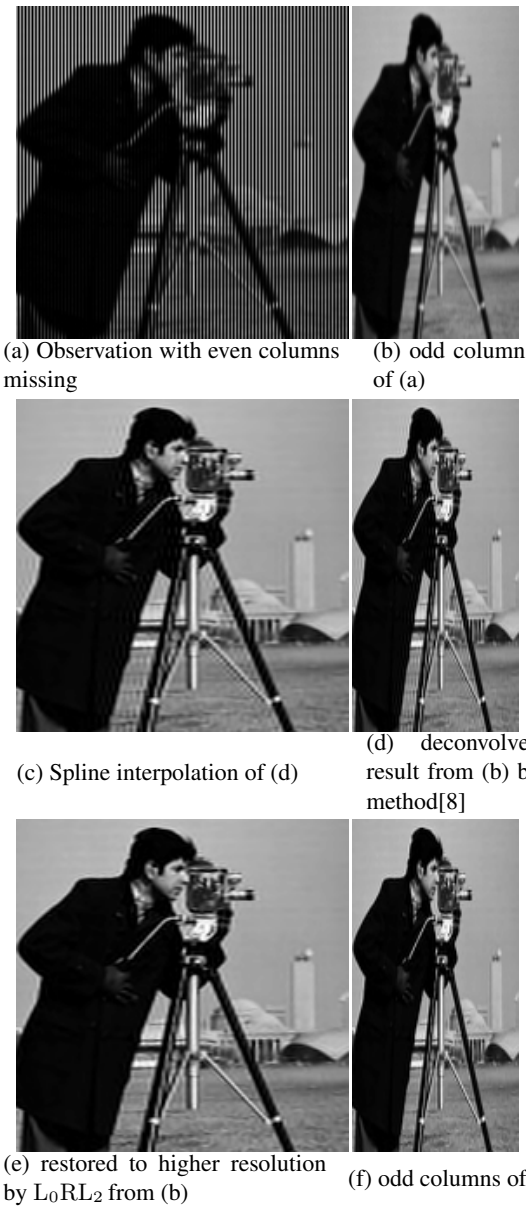
Figure 2(d) shows the result of using L_0RL_2 to deconvolve the subsampled blurred and noisy image in fig. 2(b), and fig. 2(c) shows the result of applying spline interpolation to fig. 2(d). In comparison, fig. 2(e) is the result of applying the proposed L_0RL_2 upsampling algorithm to fig. 2(b), while fig. 2(f) shows just the odd columns of this restored result for comparison with the subfigures above it. We applied 100 iterations of the L_0RL_2 algorithm to get these results, although many fewer iterations would have given almost equivalent images. We observe that (e) and (f) have improved visual metrics (sharper edges and smoother background) over (c) and (d). We use ISNR to quantify the improvement:

$$\mathrm{ISNR}(\mathbf{x}^{(n)}) = 10\log_{10}(\tfrac{\|\mathbf{y}-\mathbf{x}\|^2}{\|\mathbf{x}^{(n)}-\mathbf{x}\|^2});$$

$$\mathrm{ISNR}(\mathbf{D}(\mathbf{x}^{(n)})) = 10\log_{10}(\tfrac{\|\bar{\mathbf{y}}-\mathbf{D}\mathbf{x}\|^2}{\|\mathbf{D}\mathbf{x}^{(n)}-\mathbf{D}\mathbf{x}\|^2});$$

and get the following ISNR(dB) measures:

	(c)	(d)	(e)	(f)
Guassian filter	3.1734	3.2720	4.2188	4.0900
Uniform blur filter	4.3641	4.5664	5.5276	5.5179



(a) Observation with even columns missing

(b) odd columns of (a)

(c) Spline interpolation of (d)

(d) deconvolved result from (b) by method[8]

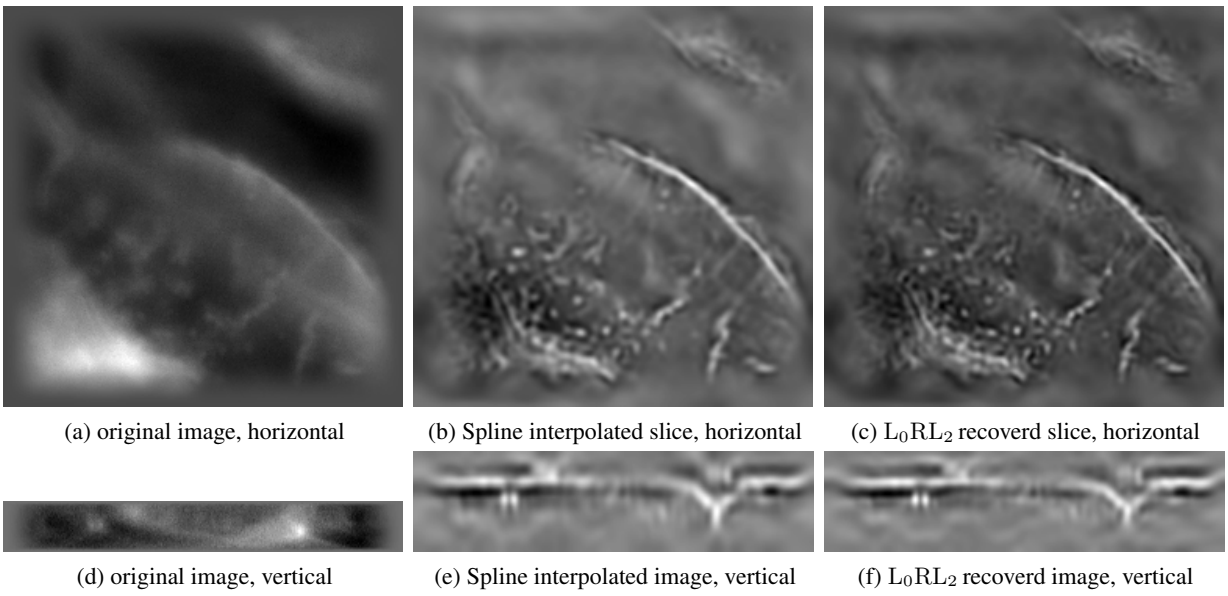
(e) restored to higher resolution by L_0RL_2 from (b)

(f) odd columns of (e)

Fig. 2. Results of L_0RL_2 on Cameramen, Gaussian filter, $[16 \times 16]$, $\sigma = 1$, BSNR = 40dB.

In the above table we also give results for an 8×8 uniform blurring kernel on the Cameraman image. We are only able to show images for the Gaussian blur kernel in Fig.2 due to space limitations.

The 3D microscopy dataset is from a wide-field fluorescence microscope experiment (Fig.4). In microscope imaging systems, the sampling rate of the vertical (inter-slice) direction is often lower than in the horizontal plane due to physical considerations. Therefore, using the L_0RL_2 algorithm to recover the missing slices and restore the dataset to equal resolution in all directions is valuable. The 3D DT $\mathbb{C}$WT representation of the dataset gives 28 directional subbands in each



(a) original image, horizontal (b) Spline interpolated slice, horizontal (c) $\mathrm{L_0RL_2}$ recoverd slice, horizontal

(d) original image, vertical (e) Spline interpolated image, vertical (f) $\mathrm{L_0RL_2}$ recoverd image, vertical

Fig. 3. Result for a 3D fluorescence microscope dataset.

scale. The coefficients in different subbands are denoised by bivariate shrinkage [4] to improve the quality of $\mathbf{A}$. The results are shown in Fig.4. Subfigure (c) shows one slice of the $\mathrm{L_0RL_2}$ recovery, while (b) shows the result obtained by interpolation of the conventionally deconvolved dataset. Subfigures (d) to (f) show vertical slices through these datasets. There is a noticeable improvement in sharpness of features in (c) and (f) over (b) and (e).

5. CONCLUSION AND DISCUSSION

By using the $\mathrm{L_0RL_2}$ algorithm to recover the image to a higher resolution, not only the deconvolved results at the observation level are improved over the conventional deconvolved result by about 0.8 dB in Cameraman, but also improved results on the higher resolution are gained. This is because the new model we use in this paper make a better use of the PSF by considering the convolution model at a higher sampling rate, rather than using the conventional discrete model at the sampling rate of the observation.

The improved orientation selectivity of DT $\mathbb{C}$WT is particularly beneficial for this missing-slice interpolation problem because it allows surfaces in diagonal directions to be emphasised by specific subbands, while noise and aliasing effects in the other directions can be suppressed by the sparsity-inducing regularization process.

6. REFERENCES

[1] I.W. Selesnick, R.G. Baraniuk, and N.G. Kingsbury, "The dual-tree complex wavelet transform," *Signal Processing Magazine, IEEE*, vol. 22, no. 6, pp. 123–151, 2005.

[2] H. Choi, J. Romberg, R. Baraniuk, and N.G. Kingsbury, "Hidden markov tree modeling of complex wavelet transforms," *ICASSP'00. Proceedings.*, vol. 1, 2000.

[3] J.K. Romberg, H. Choi, and R.G. Baraniuk, "Bayesian tree-structured image modeling using wavelet-domain hidden Markov models," *IEEE Transactions on Image Processing*, vol. 10, no. 7, pp. 1056–1068, 2001.

[4] L. Sendur and I.W. Selesnick, "Bivariate shrinkage functions for wavelet-based denoising exploiting interscale dependency," *IEEE Transactions on Signal Processing*, vol. 50, no. 11, pp. 2744–2756, 2002.

[5] S.G. Mallat, *A wavelet tour of signal processing*, Academic Prentice-Hall, Inc. Upper Saddle River, NJ, USA, 1999.

[6] R.G. Baraniuk, "Compressive sensing," *IEEE Signal Processing Magazine*, vol. 24, no. 4, pp. 118, 2007.

[7] R.G. Baraniuk, V. Cevher, M.F. Duarte, and C. Hegde, "Model-based compressive sensing," *IEEE Transactions on Information Theory*, 2008, submitted
2008.

[8] Y. Zhang and N. Kingsbury, "A Bayesian wavelet-based multidimensional deconvolution with sub-band emphasis," in *EMBS 2008. 30th Annual International Conference of the IEEE*, 2008, pp. 3024–3027.

[9] Y. Zhang and N. Kingsbury, "Image deconvolution using a Gaussian Scale Mixtures model to approximate the wavelet sparseness constraint," in *ICASSP 2009. IEEE International Conference on*, April 2009, pp. 681–684.

[10] L. He and L. Carin, "Exploiting structure in wavelet-based Bayesian compressive sensing," *IEEE Transactions on Signal Processing*, vol. 57, pp. 3488–3497, 2009.

[11] C. Vonesch and M. Unser, "A fast thresholded Landweber algorithm for wavelet-regularized multidimensional deconvolution," *IEEE Transactions on Image Processing*, vol. 17, no. 4, pp. 539-549, 2008.