SPARSITY FINE TUNING IN WAVELET DOMAIN WITH APPLICATION TO COMPRESSIVE IMAGE RECONSTRUCTION

Weisheng Dong[a], Xiaolin Wu[b,c], and Guangming Shi[a]

[a]School of Electronic Engineering, Xidian University, Xi'an, 710071, China
[b] School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, China
[c]Dept. of Electrical and Computer Engineering,
McMaster University, Hamilton, ON L8S 4K1, Canada

ABSTRACT

In compressive sensing, wavelet space is widely used to generate sparse signal (image signal in particular) representations. In this work, we propose a novel approach of statistical context modeling to increase the level of sparsity of wavelet image representations. It is shown, contrary to a widely held assumption, that high-frequency wavelet coefficients have non-zero mean distributions if conditioned on local image structures. Removing this bias can make wavelet image representations sparser, i.e., having a greater number of zero and close-to-zero coefficients. The resulting unbiased probability models can significantly improve the performance of existing wavelet-based compressive image reconstruction methods in both PSNR and visual quality.

Index Terms— Compresses sensing, wavelet-based sparse image representation, structured sparsity.

1. INTRODUCTION

The compressive sensing (CS) theory [1] allows the reconstruction of a sparse signal $\boldsymbol{x}$ from far fewer linear measurements than the number of samples required by the Nyquist sampling methods. To fulfill the promise of CS, a flurry of research activities have been devoted to a new image acquisition paradigm: taking a relatively small set of random measurements on a scene and then recovering the scene by computational compressive image reconstruction (CIR) [2, 3].

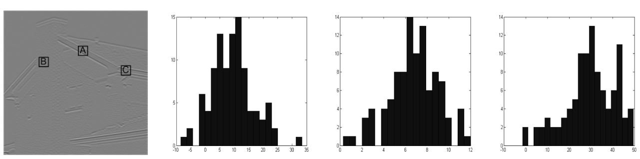
As powerful and pleasing as it is in theory, one operational hurdle of CS however requires some ingenuity to clear, that is, the determination of a reconstruction space $\boldsymbol{\Psi}$ in which the signal $\boldsymbol{x}$ has, ideally, the highest degree of sparsity possible. This is important because the sparsity level K of $\boldsymbol{x}$ under $\boldsymbol{\Psi}$ determines the fidelity of CIR. Common choices of sparsity spaces in CIR approaches are those of 2D transforms, such as discrete cosine transform (DCT), discrete wavelet transform (DWT), KL transform (KLT), etc. But these signal-independent transform spaces are adopted, to quite an extent, because of their analytical amenability and past popularity; they in their original form do not necessarily offer the sparsest representations of a given natural image.

One way of compensating for the suboptimality of reconstruction spaces is to exploit the dependencies between sparse coefficients, known as the approach of so-called structural sparsity [4, 5, 6]. The model-based CS theory developed in [4] exploits the block sparsity and tree structures of wavelet coefficients. In [5], group-sparsity regularization was proposed to exploit the tree-structured correlations of the wavelet coefficients. Bayesian CIR approaches employing sophisticated prior probability models have also been developed [6] to exploit the prior of the wavelet tree structures.

In this paper, we propose a sparsity fine tuning approach to refine the image representation in a transform space in a way such that it becomes more sparse to benefit CIR. Specifically, we develop new techniques to increase the degree of sparsity of wavelet representations. Our key observation is that though the wavelet coefficients in a high-frequency subband are zero-mean, in regions of high activities (e.g., edges and textures) the distribution of wavelet coefficients often exhibits a significant bias away from zero. Different from previous zero-mean statistical modeling of wavelet coefficients, we propose to estimate the expectations of each wavelet coefficient conditioned on local image structure. The estimated expectations are then used to remove the bias of the wavelet distributions. The maximal a posterior (MAP) estimator with the unbiased probability density functions (PDFs) are then proposed for CIR. An efficient iterative shrinkage algorithm is used to solve the resulting minimization problem. Experimental results show that the proposed CIR method can significantly improve the wavelet-based CIR and outperform most of existing CIR methods in both PSNR and visual quality.

2. SPARSITY FINE TUNING VIA CONTEXT MODELING

Wavelet space is often chosen to generate sparse image representations in CIR and other sparsity-based restoration tasks, because the majority of wavelet coefficients in high-frequency subbands are zero or very close to zero. It is commonly assumed that high-frequency wavelet coefficients obey a Laplacian or generalized Gaussian distribution of zero mean. This assumption only holds for the prior probability of wavelet coefficients in a high-frequency subband without any context. However, because image signals are not stationary in wavelet domain, if conditioning a wavelet coefficient α_i, where i indexes the location in the frequency-spatial wavelet domain, on a context $\boldsymbol{C}_i \in R^K$ of K neighboring coefficients of α_i, then the conditional PDF $P(\alpha_i|\boldsymbol{C}_i)$ generally exhibits some bias with non-zero mean. To illustrate this, we plot in Fig. 1 the conditional sample histograms in the LH wavelet subband of image *House*. It can be observed that conditional histograms differ significantly from one the other and have peaks quite far away from the origin. This property was long known to image coding researchers; they made coding gains on α_i by driving entropy coder with an estimated PDF $P(\alpha_i|\boldsymbol{C}_i)$, a process called context modeling [7]. Much like context modeling for entropy coding, we can make a wavelet image representation sparser by estimating and removing the conditional expectation $\mu_i = E\{P(\alpha_i|\boldsymbol{C}_i)\}$ from α_i. Indeed, as the resulting conditional PDF $P(\alpha_i - \mu_i|\boldsymbol{C}_i)$ is centered at zero, $(\alpha_i - \mu_i)$ is more likely to be zero or very small than α_i, meaning that the bias-removed wavelet representation becomes sparser.



(a) (b) (c) (d)

Fig. 1. (a): part of the LH subband of OWT of *House* image in the 1st-level decomposition; (b)-(d): Histograms of coefficients conditioned on contexts A, B, C, the estimated means are: (A) $\mu = 9.3$, (B) $\mu = 6.76$, (C) $\mu = 29.69$.

To accurately estimate the PDF of wavelet coefficients, we use the overcomplete wavelet transform (OWT). For capturing all inter and intra subband correlations, the context $\boldsymbol{C}_i$ should contain the coefficients near spatial position i in all subbands. However, this yields a context vector $\boldsymbol{C}_i$ of high dimensions, running the risk of context dilution(curse of dimensionality) when estimating $P(\alpha_i|\boldsymbol{C}_i)$. To prevent this problem, we simply use a image patch of size $\sqrt{K} \times \sqrt{K}$ centered at spatial position i as the context. Also, to simplify the estimation task we use a parametric model (e.g., Laplacian) for $P(\alpha_i|\boldsymbol{C}_i)$. Then the problem is reduced to the estimation of the mean μ and variance σ^2 of $P(\alpha_i|\boldsymbol{C}_i)$. To estimate μ and σ^2, we need to collect a set $S_i(\boldsymbol{C}_i)$ of samples drawn from a spatial context sufficiently close to $\boldsymbol{C}_i$. Here, we collect the samples whose contexts are within the first L closest contexts to $\boldsymbol{C}_i$.

3. THE PROPOSED COMPRESSIVE IMAGE RECONSTRUCTION METHOD

In this section we develop a Maximal a Posterior (MAP) estimator based on the above proposed unbiased Laplacian model and construct a bias-removed ℓ_1 sparse model for CIR. Given $\boldsymbol{y} = \boldsymbol{\Phi x}$, where $\boldsymbol{\Phi}$ is the measurement matrix, we jointly estimate the wavelet coefficients $\boldsymbol{\alpha}$ as well as the conditional expectations $\boldsymbol{\mu}$ by maximizing the posterior $P(\boldsymbol{\alpha}, \boldsymbol{\mu}|\boldsymbol{y})$, i.e.,

$$(\boldsymbol{\alpha}^*, \boldsymbol{\mu}^*) = \underset{\boldsymbol{\alpha},\boldsymbol{\mu}}{\operatorname{argmax}} \log P(\boldsymbol{y}|\boldsymbol{\alpha}, \boldsymbol{\mu}) + \log P(\boldsymbol{\alpha}, \boldsymbol{\mu}), \quad (1)$$

where $P(\boldsymbol{y}|\boldsymbol{\alpha})$ is the Gaussian likelihood term, modeled as

$$P(\boldsymbol{y}|\boldsymbol{\alpha}) = \frac{1}{\sqrt{2\pi}\sigma_n} \exp(-||\boldsymbol{y} - \boldsymbol{\Phi\Psi\alpha}||_2^2/(2\sigma_n^2)), \quad (2)$$

where σ_n denotes the standard deviation of the additive Gaussian noise. The prior model $P(\boldsymbol{\alpha}, \boldsymbol{\mu})$ can be written as

$$\begin{aligned} P(\boldsymbol{\alpha}, \boldsymbol{\mu}) &= \prod_i P(\alpha_i, \mu_i|\boldsymbol{C}_i) \\ &= \prod_i P(\alpha_i|\boldsymbol{C}_i, \mu_i) P(\mu_i|\boldsymbol{C}_i). \end{aligned} \quad (3)$$

The conditional prior probability model $P(\alpha_i|\boldsymbol{C}_i, \mu_i)$ is given by

$$P(\alpha_i|\boldsymbol{C}_i, \mu_i) = \frac{1}{\sqrt{2}\sigma_i} \exp(-|\alpha_i - \mu_i|/\sigma_i), \quad (4)$$

where σ_i denotes the standard deviation of α_i.

The expectations μ_i can be estimated from $S_i(\boldsymbol{C}_i)$ as a weighted average of the samples, i.e., $\hat{\mu}_i = \sum_{l \in S_i} w_l \alpha_l$. The weights w_l is computed as $w_i = \frac{1}{z} \exp(-||\boldsymbol{C}_l - \boldsymbol{C}_i||/h)$, where z denotes the normalization factor and h is a predefined constant. In general, α_l and $\boldsymbol{C}_l$ are not available. In practice, they can be obtained from initial estimate of the original image $\boldsymbol{x}$. Due to the estimation errors of w_l and α_l, $\hat{\mu}_i$ may not be accurate enough. Thus, we model $\hat{\mu}_i$ as $\hat{\mu}_i = \mu_i + e_i$, where e_i denotes the estimation error, which is assumed to be the additive Gaussian noise. Thus, $P(\mu_i|\boldsymbol{C}_i)$ can be modeled as

$$P(\mu_i|\boldsymbol{C}_i) = \frac{1}{\sqrt{2\pi}\sigma_e} \exp(-\frac{1}{2\sigma_e^2}||\mu_i - \sum_{l \in S_i} w_l \alpha_l||_2^2), \quad (5)$$

where σ_e^2 denotes the variance of e_i. By substituting Eqs.(4) and (5) into Eq.(3), we obtain the prior model $P(\boldsymbol{\alpha}, \boldsymbol{\mu})$ as

$$P(\boldsymbol{\alpha}, \boldsymbol{\mu}) = \prod_i c \cdot \exp(-\frac{1}{\sigma_i}|\alpha_i - \mu_i| - \frac{1}{2\sigma_e^2}||\mu_i - \sum_{l \in S_i} w_l \alpha_l||_2^2). \quad (6)$$

where c denotes the constant.

By substituting Eqs. (2) and (6) into Eq.(1), we obtain

$$\begin{aligned} (\boldsymbol{\alpha}, \boldsymbol{\mu}) = \underset{\boldsymbol{\alpha},\boldsymbol{\mu}}{\operatorname{argmin}} ||\boldsymbol{y} - \boldsymbol{\Phi\Psi\alpha}||_2^2 + \sum_i \lambda_i |\alpha_i - \mu_i| \\ + \eta \sum_i ||\mu_i - \sum_{l \in S_i} w_l \alpha_l||_2^2, \end{aligned} \quad (7)$$

where $\lambda_i = 2\sqrt{2}\sigma_n^2/\sigma_i$ and $\eta = \sigma_n^2/\sigma_e^2$. Compared to the conventional ℓ_1 sparse model, the above proposed bias-removed ℓ_1 sparse model enjoys two advantages. First, the nonzero means μ_i are used to reduce the magnitudes of those nonzero wavelet coefficients, and thus significantly increase the sparsity of the wavelet representation. Second, the regularization parameters λ_i are locally computed using σ_i that can be estimated from S_i. For expression convenience, we rewrite Eq.(7) as

$$(\boldsymbol{\alpha}, \boldsymbol{\mu}) = \underset{\boldsymbol{\alpha},\boldsymbol{\mu}}{\operatorname{argmin}} ||\boldsymbol{y}-\boldsymbol{\Phi\Psi\alpha}||_2^2+||\boldsymbol{\Lambda}(\boldsymbol{\alpha}-\boldsymbol{\mu})||_1+\eta||\boldsymbol{\mu}-\boldsymbol{W\alpha}||_2^2, \tag{8}$$

where $\boldsymbol{\Lambda}$ is a diagonal weighting matrix whose diagonal elements are λ_i, and the matrix $\boldsymbol{W}$ is

$$\boldsymbol{W}(i,j) = \begin{cases} w_{i,l}, \text{ if } l \in S_i, \\ 0, \text{ otherwise.} \end{cases} \tag{9}$$

4. OPTIMIZATION ALGORITHM

The proposed objective function can be solved by alternatively optimizing $\boldsymbol{\alpha}$ and $\boldsymbol{\mu}$. With an initial estimate of $\boldsymbol{\mu}$, $\boldsymbol{\alpha}$ can be solved by minimizing

$$\boldsymbol{\alpha} = \underset{\boldsymbol{\alpha}}{\operatorname{argmin}} ||\boldsymbol{y}-\boldsymbol{\Phi\Psi\alpha}||_2^2+||\boldsymbol{\Lambda}(\boldsymbol{\alpha}-\boldsymbol{\mu})||_1+\eta||\boldsymbol{\mu}-\boldsymbol{W\alpha}||_2^2, \tag{10}$$

which can be rewritten as

$$\boldsymbol{\alpha} = \underset{\boldsymbol{\alpha}}{\operatorname{argmin}} ||\tilde{\boldsymbol{y}} - \boldsymbol{A\alpha}||_2^2 + ||\boldsymbol{\Lambda}(\boldsymbol{\alpha} - \boldsymbol{\mu})||_1, \tag{11}$$

where $\tilde{\boldsymbol{y}}$ and $\boldsymbol{A}$ are defined as

$$\tilde{\boldsymbol{y}} = \begin{bmatrix} \boldsymbol{y} \\ \eta\boldsymbol{\mu} \end{bmatrix}, \quad \boldsymbol{A} = \begin{bmatrix} \boldsymbol{\Phi\Psi} \\ \eta\boldsymbol{W} \end{bmatrix}. \tag{12}$$

Eq.(11) is a weighted ℓ_1 sparse coding problem, which can be efficiently solved by the iterative shrinkage algorithm [8]. The shrinkage function for updating $\boldsymbol{\alpha}$ with fixed $\boldsymbol{\mu}^{(k)}$ can be derived as

$$\alpha_i^{(k+1)} = \boldsymbol{S}_{\tau_1}(v_i^{(k)} - \mu_i^{(k)}) + \mu_i^{(k)}, \tag{13}$$

where $\boldsymbol{S}_{\tau_1}(\cdot)$ denotes the soft thresholding operator and

$$\boldsymbol{v}^{(k)} = \frac{1}{c}\boldsymbol{A}^\top(\tilde{\boldsymbol{y}} - \boldsymbol{A\alpha}^{(k)}) + \boldsymbol{\alpha}^{(k)} \tag{14}$$

and $\tau_1 = \frac{\lambda_i}{c}$, where c is chosen such that $c > ||\boldsymbol{A}^\top\boldsymbol{A}||_2$. After solving for $\boldsymbol{\alpha}$, an estimate of the original image $\boldsymbol{x}$ can be obtained by inverse OWT as $\boldsymbol{x}^{(k+1)} = \boldsymbol{\Psi\alpha}^{(k+1)}$.

For a fixed $\boldsymbol{\alpha}^{(k+1)}$, $\boldsymbol{\mu}$ can be updated by solving

$$\boldsymbol{\mu} = \underset{\boldsymbol{\mu}}{\operatorname{argmin}} ||\boldsymbol{W\alpha}^{(k+1)}-\boldsymbol{\mu}||_2^2+\frac{1}{\eta}||\boldsymbol{\Lambda}(\boldsymbol{\alpha}^{(k+1)}-\boldsymbol{\mu})||_1, \tag{15}$$

which can be solved in a closed-form, as

$$\mu_i^{(k+1)} = \boldsymbol{S}_{\tau_2}(\boldsymbol{W\alpha}^{(k+1)} - \alpha_i^{(k+1)}) + \alpha_i^{(k+1)}, \tag{16}$$

where $\tau_2 = \frac{\lambda_i}{\eta}$. The overall algorithm for solving Eq.(8) is summarized in **Algorithm 1**.

Algorithm 1. CS via unbiased Laplacian model

• Initialization:

- Obtain an initially recovered image $\hat{\boldsymbol{x}}^{(0)}$ using a standard WSR-CS recovery method;
- Form the sample set $S_i(\boldsymbol{C}_i)$ via context modeling for α_i using $\hat{\boldsymbol{x}}^{(0)}$;

• Outer loop: for $j = 1, 2, ..., J$

- Inner loop (Solving Eq.(10) for $\boldsymbol{\alpha}$): for $k = 1, 2, \cdots, K$
  - Compute $\boldsymbol{v}^{(k)}$ via Eq.(14);
  - Compute $\boldsymbol{\alpha}^{(k+1)}$ via Eq.(13);
- Update $\boldsymbol{\mu}^{(j+1)}$ via Eq.(16)
- Image update: $\boldsymbol{x}^{(j+1)} = \boldsymbol{\Psi\alpha}^{(j+1)}$ via inverse OWT;
- If $mod(j, J_0) = 0$, update $S_i(\boldsymbol{C}_i)$, $\boldsymbol{W}$ and λ_i with the improved estimates of $\boldsymbol{x}$ and $\boldsymbol{\alpha}$, respectively.

In **Algorithm 1**, we update $S_i(\boldsymbol{C}_i)$, $\boldsymbol{W}$ and $\lambda_i = 2\sigma_n^2/\sigma_i$ with the improved estimates of $\boldsymbol{x}$ and $\boldsymbol{\alpha}$ in every J_0 iterations to save computational complexity.

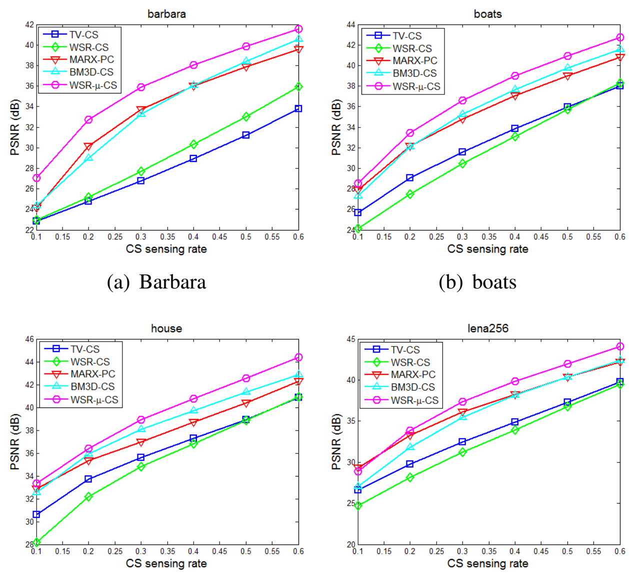
5. EXPERIMENTAL RESULTS

In this section we evaluate the performance of the proposed CIR algorithm. In our implementation, the CDF 9/7 wavelet was used and image patches of size 7×7 were used as the context. Without loss of generality, the CS measurements were generated by randomly sampling the Fourier coefficients of test images. We also compared the proposed method with the TV method of the well-known l_1-Magic software [9], the model-assisted CIR method (denoted as MARX-PC) [10], and the well-known BM3D-based CIR method (denoted as BM3D-CS) [11]. Note that the MARX-PC and BM3D-CS methods are among the state-of-the-art CIR methods.

Fig. 2 presents the PSNR curves of the tested methods, in which the number of CS measurements M is given as the percentage of the total number of pixels N, and the labels WSR-CS and WSR-μ-CS denote the CIR method using zero-mean Laplace PDF and the proposed method, respectively. As shown, the proposed WSR-μ-CS method significantly outperforms the WSR-CS method. The WSR-μ-CS method also outperforms other competing methods. The PSNR gains over the BM3D-CS method of [11] that is ranked the second in the comparison group can be up to 3.08 dB. Parts of the reconstructed images by the tested methods are shown in Figs. 3-4, which demonstrates the superiority of the proposed method over other methods. The former reproduces much sharper and cleaner image details than others.

6. ACKNOWLEDGEMENT

This work was supported in part by the Natural Science Foundation of China under Grant 61100154, Grant 61227004, Grant 61331014, Grant 61372131 and Grant 61372071, the Research Fund for the Doctoral Program of Higher Education (No. 20130203130001), and the 111 Project: No. B07048.



(a) Barbara (b) boats

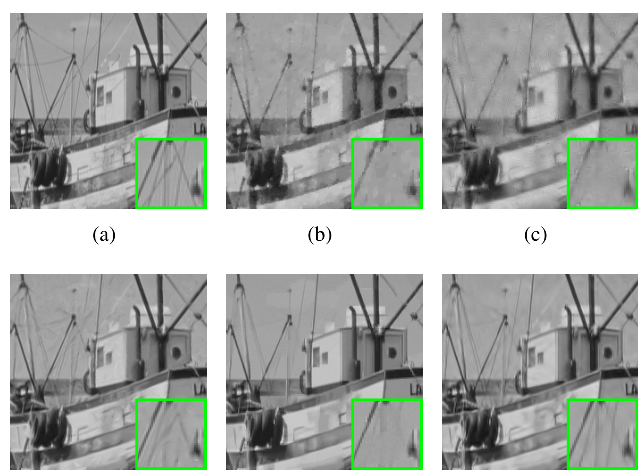
(c) house (d) lena256

Fig. 2. PSNR curves of the reconstructed images by different CIR methods.

(a) (b) (c)

(d) (e) (f)

Fig. 3. CS recovered *Barbara* images with $0.1N$ measurements. (a) Original image; (b) TV (22.79 dB); (c) WSR-CS (22.99 dB); (d) MARX-PC (24.11 dB); (e) BM3D-CS (24.34 dB); (f) Proposed WSR-μ-CS (**27.06** dB).



(a) (b) (c)

(d) (e) (f)

Fig. 4. CS recovered *Boats* images with $0.2N$ measurements. (a) Original image; (b) TV (29.01 dB); (c) WSR (27.49 dB); (d) MARX-PC (32.12 dB); (e) BM3D-CS (32.09 dB); (f) Proposed WSR-μ-CS (**33.48** dB).

7. REFERENCES

[1] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, Sep. 2006.

[2] M. Duarte, M. Davenport, D. Takbar, J. Laska, T. Sun, K. Kelly, and R. Baraniuk, "Single-pixel imaging via compressive sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 21–30, 2008.

[3] A. A. Wagadarikar, N. P. Pitsianis, X. Sun, and D. J. Brady, "Video rate spectral imaging using a coded aperture snapshot spectral imager," *Opt. Express*, vol. 17, pp. 6368–6388, 2009.

[4] R. Baraniuk, V. Cevher, M. Duarte, and C. Hegde, "Model-based compressive sensing," *IEEE Transactions on Information Theory*, vol. 56, no. 4, pp. 1982–2001, 2010.

[5] N. S. Rao, R. D. Nowak, S. J. Wright, and N. G. Kingsbury, "Convex approachs to model wavelet sparsity

patterns," in *Proceedings of IEEE Int. Conf. on Image Processing*, pp. 1917–1920.

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

[7] X. Wu, "High-order context modeling and embedded conditional entropy coding of wavelet coefficients for image compression," in *Conference Record of the Thirty-First Asilomar Conference on Signals, Systems and Computers*, 1997, vol. 2.

[8] I. Daubechies, M. Defriese, and C. DeMol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Commun. Pure Appl. Math.*, vol. 57, no. 11, pp. 1413–1457, 2004.

[9] `http://www.acm.caltech.edu./l1magic`, ,"
.

[10] X. Wu, W. Dong, X. Zhang, and G. Shi, "Model-assisted adaptive recovery of compressed sensing with imaging applications," *IEEE Trans. on Image Processing*, vol. 21, no. 2, pp. 451–458, Feb. 2012.

[11] K. Egiazarian, A. Foi, and V. Katkovnik, "Compressed sensing image reconstruction via recursive spatially adaptive filtering," in *Proceedings of IEEE Int. Conf. on Image Processing*, 2007, vol. 1.