Compressed Sensing Recovery via Collaborative Sparsity

Jian Zhang[*], Debin Zhao[*], Chen Zhao[+], Ruiqin Xiong[+], Siwei Ma[+], Wen Gao[+]

[*]School of Computer Science & Technology, Harbin Institute of Technology, China
[+]School of Electronic Engineering & Computer Science, Peking University, China
{jzhangcs, dbzhao}hit.edu.cn, {zhaochen, rqxiong, swma, wgao}pku.edu.cn

Abstract

Compressed Sensing (CS) has drawn quite an amount of attention as a joint sampling and compression approach. Its theory shows that a signal can be decoded from many fewer measurements than suggested by the Nyquist sampling theory, when the signal is sparse in some domain. So one of the most significant challenges in CS is to seek a domain where a signal can exhibit a high degree of sparsity and hence be recovered faithfully. Most of conventional CS recovery approaches, however, exploited a set of fixed bases (e.g. DCT, wavelet and gradient domain) for the entirety of a signal, which are irrespective of the nonstationarity of natural signals and cannot achieve high enough degree of sparsity, thus resulting in poor rate-distortion performance. In this paper, we propose a new framework for compressed sensing recovery via collaborative sparsity (RCoS), which enforces local two-dimensional sparsity and nonlocal three-dimensional sparsity simultaneously in an adaptive hybrid space-transform domain, thus substantially utilizing intrinsic sparsities of natural images and greatly confining the CS solution space. In addition, an efficient augmented Lagrangian based technique is developed to solve the above optimization problem. Experimental results on a wide range of natural images are presented to demonstrate the efficacy of the new CS recovery strategy.

1. Introduction

The recent development of Compressed Sensing (CS) theory has drawn quite an amount of attention as an alternative to the current methodology of sampling followed by compression [1–3]. By exploiting the redundancy existed in a signal, CS conducts sampling and compression at the same time. From many fewer acquired measurements than suggested by the Nyquist sampling theory, CS theory demonstrates that, a signal can be reconstructed with high probability when it exhibits sparsity in some domain. Specifically, a signal $\boldsymbol{u}$ of size N is said to be sparse in domain Ψ, if its transform coefficients are mostly zeros, or nearly sparse if the dominant portion of coefficients are either zeros or very close to zeros. The sparsity of $\boldsymbol{u}$ in Ψ is quantified by the number of significant elements within the coefficients vector $\Psi\boldsymbol{u}$. The signal can be perfectly recovered from observations with high probability.

Given M linear measurements, the CS recovery of $\boldsymbol{u}$ from $\boldsymbol{b}$ is formulated as the following constrained optimization problem:

$$\min_{\boldsymbol{u}} \left\|\Psi\boldsymbol{u}\right\|_p \quad \text{s.t.} \quad \boldsymbol{b} = \boldsymbol{A}\boldsymbol{u}\,, \tag{1}$$

where $\boldsymbol{A}$ represents random projections. p is usually set to 1 or 0, characterizing the sparsity of the vector $\Psi\boldsymbol{u}$. $\|*\|_1$ is l^1 norm, adding all the absolute values of the entries in a vector, while $\|*\|_0$ is l^0 norm, counting the nonzero entries of a vector.

An attractive strength of CS-based compression is that the encoder is made signal-independent and computationally inexpensive at the cost of high decoder complexity, that is, simple encoder and complex decoder, which quite resembles distributed source coding in spirit. More specifically, in the encoding process, the same random projection can be conducted on all input signals, which is non-adaptive and irrespective of any differences in their structures. It is up to the decoder to solve a large scale optimization problem to recover the randomly sampled signal in a domain where the signal exhibits sparsity. This asymmetric design is severely desirable in some image processing applications when the data acquisition devices must be simple (e.g. inexpensive resource-deprived sensors), or when oversampling can harm the object being captured (e.g. X-ray imaging) [5].

In CS theory, the sparsity degree of a signal plays a significant role in recovery. The higher degree of a signal, the higher recovery quality it will have. So, seeking a domain in which the signal has a high degree of sparsity is one of the main challenges which CS recovery should face. Since natural signals such as images are typically nonstationary, there exists no universal domain in which all parts of the signals are sparse. The most current CS recovery methods explored a set of fixed domains (e.g. DCT, wavelet and gradient domain) [16, 17, 8], therefore are signal-independent or not adaptive, resulting in poor rate-distortion performance compared to the conventional coding techniques.

To rectify the problem, Wu *et al.* [5] proposed a model-guided adaptive recovery of compressed sensing (MARX) via utilizing piecewise autoregressive model to adapt to the changing second order statistics of natural images. Many recent works incorporated additional prior knowledge about transform coefficients (statistical dependencies, structure, etc.) into the CS recovery framework, such as Gaussian scale mixtures (GSM) models [9], tree-structured wavelet [4], tree-structured DCT (TSDCT) [7]. Additionally, in [6], a projection-driven CS recovery coupled with block-based random image sampling is developed, which aims to encourage sparsity in the domain of directional transforms.

Considering the fact that the natural image signal is non-stationary and its sparse domain varies spatially, in this paper, we first establish a new sparsity measure, called collaborative sparsity measure (CoSM), and then a novel strategy for compressed sensing recovery via collaborative sparsity (RCoS) is proposed. The collaborative sparsity enforces local two-dimensional sparsity and nonlocal three-dimensional sparsity simultaneously, which offers a powerful mechanism of characterizing the structured sparsities of a natural image and enables a natural image to be highly sparse in an adaptive hybrid space-transform domain. To make RCoS tractable and robust, an augmented Lagrangian based technique is developed to solve the above severely underdetermined inverse problem efficiently. Extensive experiments on a wide range of CS-acquired images manifest that RCoS is able to increase recovery quality by a large margin compared with the conventional CS recovery methods or require many fewer measurements for a desired reconstruction quality.

The remainder of the paper is organized as follows. Section 2 introduces the concept of collaborative sparsity and shows how this type of sparsity is incorporated into the framework of CS recovery. Section 3 gives the implementation details of RCoS. Experimental results are reported in Section 4. In Section 5, we conclude this paper.

2. Collaborative Sparsity in a Hybrid Space-Transform Domain

In this section, a generic collaborative sparsity measure (CoSM) for high fidelity of CS recovery is established in an adaptive hybrid space-transform domain by merging two complementary sparsities– local two-dimensional sparsity in space domain Ψ_{L2D} and nonlocal three-dimensional sparsity in transform domain Ψ_{N3D}. The former type of sparsity describes the piecewise smoothness, while the latter one depicts the self-similarity of natural images, retaining the sharpness and edges effectively. This collaborative sparsity measure can combine the best of the both worlds, and significantly improve the reconstruction quality of CS-acquired images. More specifically, the collaborative sparsity recovery of CS is formulated in the following:

$$\min_{\boldsymbol{u}} \left\|\Psi_{L2D}\boldsymbol{u}\right\|_p + \alpha\left\|\Psi_{N3D}\boldsymbol{u}\right\|_q \quad \text{s.t.} \quad \boldsymbol{b} = \boldsymbol{A}\boldsymbol{u}\,, \tag{2}$$

where p and q are usually set to values from the interval $[0,1]$, and α is a regularization parameter, which controls the trade-off between two competing sparsity terms.

2.1. Local two-dimensional sparsity in space domain

To characterize the smoothness of images, there exist many models. From the view of statistics, the image is preferred when its responses for a set of filters are as small as possible. In practice, the widely-used filters are vertical and horizontal finite difference operators, denoted by $\mathcal{D}_{\mathrm{v}}$ and $\mathcal{D}_{\mathrm{h}}$, respectively. Without loss of generality, let $\Psi_{L2D} = \left[\mathcal{D}_{\mathrm{v}}; \mathcal{D}_{\mathrm{h}}\right]$ and set p to be 1 in Problem (2) to achieve the local two-dimensional sparsity in space domain, expressed as

$$\left\|\Psi_{L2D}\boldsymbol{u}\right\|_1 = \left\|\mathcal{D}_{\mathrm{v}}\boldsymbol{u}\right\|_1 + \left\|\mathcal{D}_{\mathrm{h}}\boldsymbol{u}\right\|_1, \tag{3}$$

which essentially underlies the fact that an image exhibits Laplacian sparsity in space domain. It can also be useful to include the second order derivatives or the more sophisticated learned filters [11] and set p to be 1/2 or 2/3 as hyper-Laplacian priors [10].

2.2. Nonlocal three-dimensional sparsity in transform domain

Motivated by the success of sparse representation [15] and self-similarity [12] in image restoration [13], we integrate them and introduce a nonlocal three-dimensional sparsity in transform domain Ψ_{N3D} for CS recovery.

Definitions on Nonlocal Three-dimensional Sparsity in Transform Domain

1: Divide the image $\boldsymbol{x}$ with size N into n overlapped blocks of size B_s and each block is denoted by $\boldsymbol{x}_k$, i.e., $k = 1,2,...,n$.

2: Define $\boldsymbol{S}_{\boldsymbol{x}_k}$ the set including the c best matched blocks to $\boldsymbol{x}_k$ in the $L \times L$ training window, that is, $\boldsymbol{S}_{\boldsymbol{x}_k} = \left\{\boldsymbol{S}_{\boldsymbol{x}_k \otimes 1}, \boldsymbol{S}_{\boldsymbol{x}_k \otimes 2}, ..., \boldsymbol{S}_{\boldsymbol{x}_k \otimes c}\right\}$.

3: For every $\boldsymbol{S}_{\boldsymbol{x}_k}$, a group is formed by stacking the blocks belonging to $\boldsymbol{S}_{\boldsymbol{x}_k}$ into a three-dimensional array, which is denoted by $\boldsymbol{Z}_{\boldsymbol{x}_k}$.

4: Denote $\boldsymbol{T}^{3D}$ the operator of a three-dimensional transform, and $\boldsymbol{T}^{3D}(\boldsymbol{Z}_{\boldsymbol{x}_k})$ the transform coefficients for $\boldsymbol{Z}_{\boldsymbol{x}_k}$ in domain Ψ_{N3D}. Let $\Theta_{\boldsymbol{x}}$ be the column vector with size $K = B_s * c * n$ built from all the $\boldsymbol{T}^{3D}(\boldsymbol{Z}_{\boldsymbol{x}_k})$ arranged in lexicographic order.

Therefore, the nonlocal three-dimensional sparsity in transform domain Ψ_{N3D} is written as

$$\left\|\Psi_{N3D}\boldsymbol{x}\right\|_0 = \left\|\Theta_{\boldsymbol{x}}\right\|_0 = \sum_{k=1}^{n}\left\|\boldsymbol{T}^{3D}(\boldsymbol{Z}_{\boldsymbol{x}_k})\right\|_0. \tag{4}$$

Similarly, the inverse operator Ω_{N3D} corresponding to Ψ_{N3D} can be defined in the following procedures. After obtaining Θ_x, split it into n groups of three-dimensional transform coefficients, which are then inverted to generate estimates for each block in the group. The block-wise estimates are returned to their original positions and the final image estimate is achieved by averaging all of the above block-wise estimates. Therefore, given Θ_x, the new estimate for $\boldsymbol{x}$ is expressed as $\hat{\boldsymbol{x}} = \Omega_{N3D}\Theta_x$.

3. CS Recovery via Collaborative Sparsity

Incorporating (3) and (4) into (2), the proposed constrained optimization problem for CS recovery via collaborative sparsity (RCoS) in an adaptive hybrid space-transform domain is formulated as:

$$\min_{\boldsymbol{u}} \left\|\mathcal{D}_{\mathrm{v}}\boldsymbol{u}\right\|_1 + \left\|\mathcal{D}_{\mathrm{h}}\boldsymbol{u}\right\|_1 + \alpha\left\|\Theta_{\boldsymbol{u}}\right\|_0 \ \text{ s.t. } \ \boldsymbol{A}\boldsymbol{u} = \boldsymbol{b}. \tag{5}$$

Note that Problem (5) is essentially non-convex and quite difficult to solve directly due to the non-differentiability and non-linearity of the collaborative sparsity term. Solving it efficiently is one of the main contributions of this paper. In this section, the implementation details of RCoS are provided.

Instead of solving (5) directly, Problem (5) is first transformed into an equivalent variant by introducing auxiliary variables $\boldsymbol{w}_{\mathrm{v}}, \boldsymbol{w}_{\mathrm{h}}$ and $\boldsymbol{x}$:

$$\min_{\boldsymbol{w}_{\mathrm{v}},\boldsymbol{w}_{\mathrm{h}},\boldsymbol{u},\boldsymbol{x}} \left\|\boldsymbol{w}_{\mathrm{v}}\right\|_1 + \left\|\boldsymbol{w}_{\mathrm{h}}\right\|_1 + \alpha\left\|\Theta_{\boldsymbol{x}}\right\|_0 \ \text{ s.t. } \ \mathcal{D}_{\mathrm{v}}\boldsymbol{u} = \boldsymbol{w}_{\mathrm{v}}, \mathcal{D}_{\mathrm{h}}\boldsymbol{u} = \boldsymbol{w}_{\mathrm{h}}, \boldsymbol{u} = \boldsymbol{x}, \boldsymbol{A}\boldsymbol{u} = \boldsymbol{b}. \tag{6}$$

The corresponding augmented Lagrangian function of (6) is

$$\begin{aligned} \mathcal{L}_{\mathcal{A}}\left(\boldsymbol{w}_{\mathrm{v}}, \boldsymbol{w}_{\mathrm{h}}, \boldsymbol{u}, \boldsymbol{x}\right) = {} & \left\|\boldsymbol{w}_{\mathrm{v}}\right\|_1 - \mathbf{v}_{\mathrm{v}}^T\left(\mathcal{D}_{\mathrm{v}}\boldsymbol{u} - \boldsymbol{w}_{\mathrm{v}}\right) + \tfrac{\beta_{\mathrm{v}}}{2}\left\|\mathcal{D}_{\mathrm{v}}\boldsymbol{u} - \boldsymbol{w}_{\mathrm{v}}\right\|_2^2 + \left\|\boldsymbol{w}_{\mathrm{h}}\right\|_1 \\ & - \mathbf{v}_{\mathrm{h}}^T\left(\mathcal{D}_{\mathrm{h}}\boldsymbol{u} - \boldsymbol{w}_{\mathrm{h}}\right) + \tfrac{\beta_{\mathrm{h}}}{2}\left\|\mathcal{D}_{\mathrm{h}}\boldsymbol{u} - \boldsymbol{w}_{\mathrm{h}}\right\|_2^2 + \alpha\left\|\Theta_{\boldsymbol{x}}\right\|_0 - \boldsymbol{\gamma}^T\left(\boldsymbol{u} - \boldsymbol{x}\right) \\ & + \tfrac{\theta}{2}\left\|\boldsymbol{u} - \boldsymbol{x}\right\|_2^2 + \tfrac{\mu}{2}\left\|\boldsymbol{A}\boldsymbol{u} - \boldsymbol{b}\right\|_2^2 - \boldsymbol{\lambda}^T\left(\boldsymbol{A}\boldsymbol{u} - \boldsymbol{b}\right) \end{aligned} \tag{7}$$

where $\beta_{\mathrm{v}}, \beta_{\mathrm{h}}, \theta, \mu$ are regularization parameters associated with quadratic penalty terms $\left\|\mathcal{D}_{\mathrm{v}}\boldsymbol{u} - \boldsymbol{w}_{\mathrm{v}}\right\|_2^2, \left\|\mathcal{D}_{\mathrm{h}}\boldsymbol{u} - \boldsymbol{w}_{\mathrm{h}}\right\|_2^2, \left\|\boldsymbol{u} - \boldsymbol{x}\right\|_2^2, \left\|\boldsymbol{A}\boldsymbol{u} - \boldsymbol{b}\right\|_2^2$, respectively.

The basic idea of the augmented Lagrangian method is to seek a saddle point of $\mathcal{L}_{\mathcal{A}}\left(\boldsymbol{w}_{\mathrm{v}}, \boldsymbol{w}_{\mathrm{h}}, \boldsymbol{u}, \boldsymbol{x}\right)$, which is also the solution of Problem (5). We utilize the augmented Lagrangian method to solve constrained Problem (6) by iteratively solving Problems (8) and (9):

$$\left(\boldsymbol{w}_{\mathrm{v},k+1}, \boldsymbol{w}_{\mathrm{h},k+1}, \boldsymbol{u}_{k+1}, \boldsymbol{x}_{k+1}\right) = \operatorname*{argmin}_{\boldsymbol{w}_{\mathrm{v}},\boldsymbol{w}_{\mathrm{h}},\boldsymbol{u},\boldsymbol{x}} \mathcal{L}_{\mathcal{A}}\left(\boldsymbol{w}_{\mathrm{v}}, \boldsymbol{w}_{\mathrm{h}}, \boldsymbol{u}, \boldsymbol{x}\right), \tag{8}$$

$$\begin{cases} \mathbf{v}_{\mathrm{v},k+1} = \mathbf{v}_{\mathrm{v},k} - \beta_{\mathrm{v}}\left(\mathcal{D}_{\mathrm{v}}\boldsymbol{u}_{k+1} - \boldsymbol{w}_{\mathrm{v},k+1}\right),\ \boldsymbol{\gamma}_{k+1} = \boldsymbol{\gamma}_k - \theta\left(\boldsymbol{u}_{k+1} - \boldsymbol{x}_{k+1}\right) \\ \mathbf{v}_{\mathrm{h},k+1} = \mathbf{v}_{\mathrm{h},k} - \beta_{\mathrm{h}}\left(\mathcal{D}_{\mathrm{h}}\boldsymbol{u}_{k+1} - \boldsymbol{w}_{\mathrm{h},k+1}\right),\ \boldsymbol{\lambda}_{k+1} = \boldsymbol{\lambda}_k - \mu\left(\boldsymbol{A}\boldsymbol{u}_{k+1} - \boldsymbol{b}\right) \end{cases}, \tag{9}$$

here, the subscript k denotes the iteration number, and $\mathbf{v}_{\mathrm{v}}, \mathbf{v}_{\mathrm{h}}, \boldsymbol{\gamma}, \boldsymbol{\lambda}$ are the Lagrangian multipliers associated with the constraints $\mathcal{D}_{\mathrm{v}}\boldsymbol{u} = \boldsymbol{w}_{\mathrm{v}}, \mathcal{D}_{\mathrm{h}}\boldsymbol{u} = \boldsymbol{w}_{\mathrm{h}}$, $\boldsymbol{u} = \boldsymbol{x}$, $\boldsymbol{A}\boldsymbol{u} = \boldsymbol{b}$, respectively.

In addition, Problem (8) is still hard to solve efficiently in a direct way due to its non-differentiability. Here, an alternating direction technique is employed, which alternatively minimizes one variable while fixing the other variables, to split Problem (8) into the

following four sub-problems. In the following, we argue that the every separated sub-problem admits an efficient solution. For simplicity, the subscript k is omitted without confusion.

3.1. w_{v} sub-problem

Given $\boldsymbol{w}_{\mathrm{h}}, \boldsymbol{u}, \boldsymbol{x}$, the optimization problem associated with $\boldsymbol{w}_{\mathrm{v}}$ can be expressed as

$$\min_{\boldsymbol{w}_{\mathrm{v}}} \mathcal{Q}_1\left(\boldsymbol{w}_{\mathrm{v}}\right)=\min_{\boldsymbol{w}_{\mathrm{v}}}\left\{\left\|\boldsymbol{w}_{\mathrm{v}}\right\|_1-\mathbf{v}_{\mathrm{v}}^T\left(\mathcal{D}_{\mathrm{v}}\boldsymbol{u}-\boldsymbol{w}_{\mathrm{v}}\right)+\frac{\beta_{\mathrm{v}}}{2}\left\|\mathcal{D}_{\mathrm{v}}\boldsymbol{u}-\boldsymbol{w}_{\mathrm{v}}\right\|_2^2\right\}. \tag{10}$$

According to [14], the closed form of (10) is formulated as

$$\tilde{\boldsymbol{w}}_{\mathrm{v}}=\mathbf{max}\left\{\left|\mathcal{D}_{\mathrm{v}}\boldsymbol{u}-\frac{\mathbf{v}_{\mathrm{v}}}{\beta_{\mathrm{v}}}\right|-\frac{1}{\beta_{\mathrm{v}}},0\right\}\cdot\mathbf{sgn}\left(\mathcal{D}_{\mathrm{v}}\boldsymbol{u}-\frac{\mathbf{v}_{\mathrm{v}}}{\beta_{\mathrm{v}}}\right), \tag{11}$$

where $\cdot$ stands for the element-wise product of two vectors.

3.2. w_{h} sub-problem

Similarly, the closed form $\boldsymbol{w}_{\mathrm{h}}$ for the optimization problem $\mathbf{min}\mathcal{Q}_2\left(\boldsymbol{w}_{\mathrm{h}}\right)$, where $\mathcal{Q}_2\left(\boldsymbol{w}_{\mathrm{h}}\right)=\left\|\boldsymbol{w}_{\mathrm{h}}\right\|_1-\mathbf{v}_{\mathrm{h}}^T\left(\mathcal{D}_{\mathrm{h}}\boldsymbol{u}-\boldsymbol{w}_{\mathrm{h}}\right)+\frac{\beta_{\mathrm{h}}}{2}\left\|\mathcal{D}_{\mathrm{h}}\boldsymbol{u}-\boldsymbol{w}_{\mathrm{h}}\right\|_2^2$, is represented as

$$\tilde{\boldsymbol{w}}_{\mathrm{h}}=\mathbf{max}\left\{\left|\mathcal{D}_{\mathrm{h}}\boldsymbol{u}-\frac{\mathbf{v}_{\mathrm{h}}}{\beta_{\mathrm{h}}}\right|-\frac{1}{\beta_{\mathrm{h}}},0\right\}\cdot\mathbf{sgn}\left(\mathcal{D}_{\mathrm{h}}\boldsymbol{u}-\frac{\mathbf{v}_{\mathrm{h}}}{\beta_{\mathrm{h}}}\right), \tag{12}$$

where $\cdot$ stands for the element-wise product of two vectors.

3.3. u sub-problem

With the aid of $\boldsymbol{w}_{\mathrm{v}}, \boldsymbol{w}_{\mathrm{h}}, \boldsymbol{x}$, the $\boldsymbol{u}$ sub-problem is equivalent to

$$\min_{\boldsymbol{u}}\mathcal{Q}_3\left(\boldsymbol{u}\right)=\min_{\boldsymbol{u}}\begin{Bmatrix}-\mathbf{v}_{\mathrm{v}}^T\left(\mathcal{D}_{\mathrm{v}}\boldsymbol{u}-\boldsymbol{w}_{\mathrm{v}}\right)+\frac{\beta_{\mathrm{v}}}{2}\left\|\mathcal{D}_{\mathrm{v}}\boldsymbol{u}-\boldsymbol{w}_{\mathrm{v}}\right\|_2^2-\mathbf{v}_{\mathrm{h}}^T\left(\mathcal{D}_{\mathrm{h}}\boldsymbol{u}-\boldsymbol{w}_{\mathrm{h}}\right)+\frac{\beta_{\mathrm{h}}}{2}\left\|\mathcal{D}_{\mathrm{h}}\boldsymbol{u}-\boldsymbol{w}_{\mathrm{h}}\right\|_2^2\\ -\boldsymbol{\gamma}^T\left(\boldsymbol{u}-\boldsymbol{x}\right)+\frac{\theta}{2}\left\|\boldsymbol{u}-\boldsymbol{x}\right\|_2^2-\boldsymbol{\lambda}^T\left(\boldsymbol{A}\boldsymbol{u}-\boldsymbol{b}\right)+\frac{\mu}{2}\left\|\boldsymbol{A}\boldsymbol{u}-\boldsymbol{b}\right\|_2^2\end{Bmatrix}. \tag{13}$$

Clearly, $\mathcal{Q}_3\left(\boldsymbol{u}\right)$ is a quadratic function and its gradient can be expressed as

$$\boldsymbol{d}\left(\mathcal{Q}_3\left(\boldsymbol{u}\right)\right)=\mathcal{D}_{\mathrm{v}}^T\left(\beta_{\mathrm{v}}\mathcal{D}_{\mathrm{v}}\boldsymbol{u}-\mathbf{v}_{\mathrm{v}}\right)+\mathcal{D}_{\mathrm{h}}^T\left(\beta_{\mathrm{h}}\mathcal{D}_{\mathrm{h}}\boldsymbol{u}-\mathbf{v}_{\mathrm{h}}\right)-\boldsymbol{\gamma}+\theta\left(\boldsymbol{u}-\boldsymbol{x}\right)+\boldsymbol{A}^T\left(\mu\left(\boldsymbol{A}\boldsymbol{u}-\boldsymbol{b}\right)-\boldsymbol{\lambda}\right). \tag{14}$$

Setting $\boldsymbol{d}\left(\mathcal{Q}_3\left(\boldsymbol{u}\right)\right)=0$ gives us the exact minimizer of Problem (12), that is,

$$\boldsymbol{u}=\left(\beta_{\mathrm{v}}\mathcal{D}_{\mathrm{v}}^T\mathcal{D}_{\mathrm{v}}+\beta_{\mathrm{h}}\mathcal{D}_{\mathrm{h}}^T\mathcal{D}_{\mathrm{h}}+\theta\boldsymbol{I}+\mu\boldsymbol{A}^T\boldsymbol{A}\right)^{+}\left(\mathcal{D}_{\mathrm{v}}^T\mathbf{v}_{\mathrm{v}}+\mathcal{D}_{\mathrm{h}}^T\mathbf{v}_{\mathrm{h}}+\boldsymbol{\gamma}+\theta\boldsymbol{x}+\boldsymbol{A}^T\boldsymbol{\lambda}+\mu\boldsymbol{A}^T\boldsymbol{b}\right), \tag{15}$$

where $\mathbf{M}^{+}$ stands for the Moore-Penrose pseudoinverse of matrix $\mathbf{M}$. Theoretically, it is ideal to accept the exact minimizer as the solution of the $\boldsymbol{u}$ sub-problem. However, computing the inverse or pseudoinverse at each iteration is too costly to implement numerically. Therefore, an iterative method is highly desirable. Here, the steepest descent method with the optimal step is used to solve Problem (13) iteratively by applying

$$\tilde{\boldsymbol{u}}=\boldsymbol{u}-\eta\boldsymbol{d}, \tag{16}$$

where $\boldsymbol{d}$ is the gradient direction of the objective function $\mathcal{Q}_3\left(\boldsymbol{u}\right)$, $\eta=\mathbf{abs}(\boldsymbol{d}^T\boldsymbol{d}/\boldsymbol{d}^T\boldsymbol{G}\boldsymbol{d})$ is the optimal step, $\boldsymbol{G}=(\beta_{\mathrm{v}}\mathcal{D}_{\mathrm{v}}^T\mathcal{D}_{\mathrm{v}}+\beta_{\mathrm{h}}\mathcal{D}_{\mathrm{h}}^T\mathcal{D}_{\mathrm{h}}+\theta\boldsymbol{I}+\mu\boldsymbol{A}^T\boldsymbol{A})$, and $\boldsymbol{I}$ is the identity matrix.

3.4. x sub-problem

Given $\boldsymbol{w}_{\mathrm{v}}, \boldsymbol{w}_{\mathrm{h}}, \boldsymbol{u}$, the $\boldsymbol{x}$ sub-problem becomes

$$\min_{\boldsymbol{x}} \mathcal{Q}_4(\boldsymbol{x}) = \min_{\boldsymbol{x}} \left\{ \alpha \left\| \Theta_{\boldsymbol{x}} \right\|_0 - \boldsymbol{\gamma}^T (\boldsymbol{u} - \boldsymbol{x}) + \frac{\theta}{2} \left\| \boldsymbol{u} - \boldsymbol{x} \right\|_2^2 \right\}. \tag{17}$$

After making a straightforward complete-the-squares procedure and omitting a constant independent of $\boldsymbol{x}$, Problem (17) can be written as

$$\min_{\boldsymbol{x}} \mathcal{Q}_4(\boldsymbol{x}) = \min_{\boldsymbol{x}} \left\{ \frac{\theta}{2} \left\| \boldsymbol{x} - \boldsymbol{r} \right\|_2^2 + \alpha \left\| \Theta_{\boldsymbol{x}} \right\|_0 \right\} = \min_{\boldsymbol{x}} \left\{ \frac{1}{2} \left\| \boldsymbol{x} - \boldsymbol{r} \right\|_2^2 + \frac{\alpha}{\theta} \sum_{k=1}^{n} \left\| \boldsymbol{T}^{3D}(\boldsymbol{Z}_{\boldsymbol{x}_k}) \right\|_0 \right\}, \tag{18}$$

where $\boldsymbol{r} = (\boldsymbol{u} - \frac{\boldsymbol{\gamma}}{\theta})$. Here, we model the elements of $\boldsymbol{x} - \boldsymbol{r}$ as random variables from a Gaussian process with zero mean and variance σ^2, which is reasonable and commonly used in practice. Under this assumption, since $\boldsymbol{x}, \boldsymbol{r} \in \mathbf{R}^N, \Theta_{\boldsymbol{x}}, \Theta_{\boldsymbol{r}} \in \mathbf{R}^K$, and the transform $\boldsymbol{T}^{3D}$ is orthogonal for every group, there exist the following two equations with very large probability (limited to 1):

$$\left\| \boldsymbol{x} - \boldsymbol{r} \right\|_2^2 \Big/ N = \sigma^2, \quad \left\| \Theta_{\boldsymbol{x}} - \Theta_{\boldsymbol{r}} \right\|_2^2 \Big/ K = \sum_{k=1}^{n} \left\| \boldsymbol{T}^{3D}(\boldsymbol{Z}_{\boldsymbol{x}_k}) - \boldsymbol{T}^{3D}(\boldsymbol{Z}_{\boldsymbol{r}_k}) \right\|_2^2 \Big/ K = \sigma^2. \tag{19}$$

Incorporating (19) into (18) leads to

$$\min_{\boldsymbol{x}} \mathcal{Q}_4(\boldsymbol{x}) = \min_{\boldsymbol{x}} \frac{1}{2} \left\| \Theta_{\boldsymbol{x}} - \Theta_{\boldsymbol{r}} \right\|_2^2 + \frac{K\alpha}{N\theta} \left\| \Theta_{\boldsymbol{x}} \right\|_0. \tag{20}$$

Owing to [14] the closed form of (20) is written as $\tilde{\Theta}_{\boldsymbol{x}} = \mathbf{hard}\left(\Theta_{\boldsymbol{r}}, \sqrt{2\tau} \right)$, where $\tau = K\alpha / N\theta$, $\mathbf{hard}(\boldsymbol{z}, a) = \mathbf{max}\left\{ \mathbf{abs}(\boldsymbol{z} - a), 0 \right\} \cdot \mathbf{sgn}(\boldsymbol{z})$ and $\cdot$ stands for the element-wise product of two vectors. Thus, the solution for the $\boldsymbol{x}$ sub-problem (17) is

$$\tilde{\boldsymbol{x}} = \Omega_{N3D} \tilde{\Theta}_{\boldsymbol{x}} = \Omega_{N3D} \left(\mathbf{hard}(\Theta_{\boldsymbol{r}}, \sqrt{2\tau}) \right). \tag{21}$$

3.5. Summary of RCoS

So far, all issues in the process of handing the sub-problems have been solved. In light of all derivations above, the complete description of RCoS for CS-acquired images has two loops, as stated below:

CS Recovery via Collaborative Sparsity

Input: The observed measurement $\boldsymbol{b}$, the measurement matrix $\boldsymbol{A}$ and $\beta_{\mathrm{v}}, \beta_{\mathrm{h}}, \theta, \mu, \alpha$.
Initialization: $\boldsymbol{u}_0 = \boldsymbol{A}^T \boldsymbol{b}$, $\boldsymbol{w}_{\mathrm{v},0} = \boldsymbol{w}_{\mathrm{h},0} = \boldsymbol{x}_0 = \mathbf{0}$, $\mathbf{v}_{\mathrm{v},0} = \mathbf{v}_{\mathrm{h},0} = \boldsymbol{\gamma}_0 = \boldsymbol{\lambda}_0 = \mathbf{0}$;
while Outer stopping criteria unsatisfied ***do***
while Inner stopping criteria unsatisfied ***do***
Solve $\boldsymbol{w}_{\mathrm{v}}$ sub-problem by computing Eq. (11);
Solve $\boldsymbol{w}_{\mathrm{h}}$ sub-problem by computing Eq. (12);
Solve $\boldsymbol{u}$ sub-problem by computing Eq. (16);
Solve $\boldsymbol{x}$ sub-problem by computing Eq. (21);
end while
Update multipliers $\mathbf{v}_{\mathrm{v}}, \mathbf{v}_{\mathrm{h}}, \boldsymbol{\gamma}, \boldsymbol{\lambda}$ by computing Eq. (9);
end while
Output: Final restored image $\tilde{\boldsymbol{u}}$.

4. Experimental Results

In this section, the experimental results are presented to evaluate the performance of the proposed RCoS. For thoroughness and fairness of our comparative study, we exploit a broad class of natural images, ranging from conventional images to biomedical images,

as shown in Figure 1. In our experiments, the CS measurements are obtained by applying a random projection matrix to the original image signal. RCoS is compared with two representative CS recovery methods in literature, i.e., tree-structured DCT (TSDCT) method [7] and total variation (TV) method [8], which deal with the image signal in the DCT domain and the gradient domain, respectively. It is worth emphasizing that total variation (TV) method is known as one of the state-of-the-art algorithms for image CS recovery. The RCoS software and more experimental results can be found at: http://idm.pku.edu.cn/staff/zhangjian/RCoS/.

Figure 1: Experimental test images. Top: *Barbara*, *Vessels*, *Leaves* and *Parrot*; Bottom: *Chest*, *Lena*, *House* and *Cameraman*.

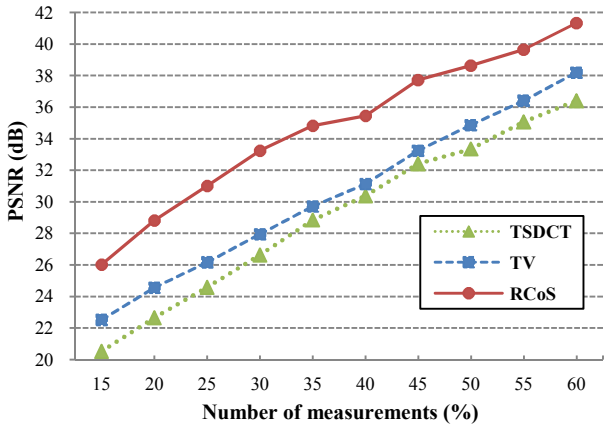


Figure 2: PSNR vs. number of CS measurements for Image *Vessels*.

The curve of PSNR versus the number of CS measurements (presented as the percentage of the total number of pixels) from 15% to 60% of Image *Vessels* for the three different approaches is plotted in Figure 2. It is obvious to see that the recovered results acquired by TV [8] are better than those by TSDCT [7]. Moreover, the performance of the proposed RCoS is consistently superior to the other two methods under the same numbers of measurements. Additionally, as show in Figure 2, for a desired reconstruction quality, such as 33.5 dB, the measurements that RCoS requires (about 30%) are much fewer than those TSDCT or TV needs (about 45%).

The PSNR comparisons for all the test images in the cases of 20% and 30% measurements are provided in Table 1. RCoS achieves the highest PSNR among the three comparative algorithms over all the cases, which can improve roughly 6 dB and 3 dB on average, compared with TSDCT and TV, respectively.

Some visual results of parts of the recovered images for the three algorithms are presented in Figures. 3~5, which verify the superiority of RCoS in preserving the image structures and fine details, showing much better visual results than the other two competing methods. It is also important to observe that the recovered results by RCoS in the case of 20% measurements are much better than those by TV or TSDCT in the case of 30% in terms of visual perception. The high performance of RCoS is attributed to the proposed adaptive hybrid space-transform domain, which offers a powerful mechanism of characterizing the structured sparsities of natural image signals.

Table 1: PSNR comparisons with different methods (dB)
(In the cases of 20% and 30% measurements)

Image	***Barbara***		***Vessels***		***Leaves***		***Parrot***	
ratio (%)	20	30	20	30	20	30	20	30
TSDCT [7]	22.87	25.44	22.64	26.62	19.97	22.81	27.92	31.55
TV [8]	24.42	26.31	24.56	27.94	23.08	26.61	29.44	32.21
RCoS	**27.67**	**31.44**	**28.81**	**33.24**	**28.18**	**31.81**	**32.33**	**34.52**
Image	***Chest***		***Lena***		***House***		***Cameraman***	
ratio (%)	20	30	20	30	20	30	20	30
TSDCT [7]	20.77	23.60	26.43	28.75	31.40	33.47	24.79	27.35
TV [8]	28.76	34.45	28.69	31.14	33.05	34.96	27.74	30.26
RCoS	**31.35**	**35.99**	**30.86**	**33.58**	**35.26**	**36.30**	**29.74**	**31.70**

5. Conclusions

In this paper, a novel sparsity measure, called collaborative sparsity measure is introduced, and a new strategy for compressed sensing recovery via collaborative sparsity (RCoS) is proposed, which efficiently characterizes the intrinsic sparsities of natural images in an adaptive hybrid space-transform domain. Extensive experiments on a wide range of CS-acquired images manifest that RCoS is able to increase recovery quality by a large margin compared with the current methods or require many fewer measurements for a desired reconstruction quality. Our work offers a fresh and successful instance to corroborate the CS theory applied for real signals (i.e., natural images).

Acknowledgement

We would like to thank the authors of [7] and [8] for kindly providing their codes. This work is supported in part by National Science Foundation (No. 60736043 and 61073083), Beijing Natural Science Foundation (No. 4112026), Specialized Research Fund for the Doctoral Program of Higher Education (SRFDP) and the Major State Basic Research Development Program of China (2009CB320905).

References

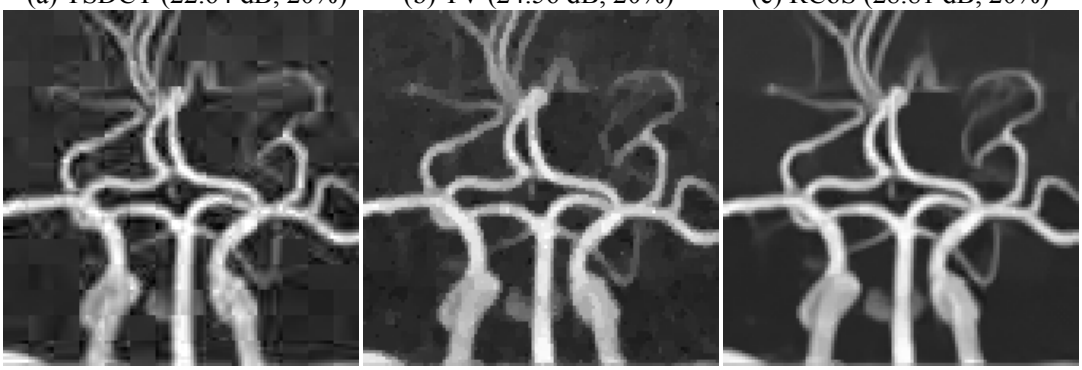
[1] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inform. Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

(a) TSDCT (21.79 dB, 20%) (b) TV (22.49 dB, 20%) (c) RCoS (26.41 dB, 20%)

(d) TSDCT (24.53 dB, 30%) (e) TV (24.67 dB, 30%) (f) RCoS (30.08 dB, 30%)

Figure 3: CS recovered portion of Image *Barbara* (20% and 30% measurements).

(a) TSDCT (22.64 dB, 20%) (b) TV (24.56 dB, 20%) (c) RCoS (28.81 dB, 20%)



(d) TSDCT (26.62 dB, 30%) (e) TV (27.94 dB, 30%) (f) RCoS (33.24 dB, 30%)

Figure 4: CS recovered Image *Vessels* (20% and 30% measurements).

(a) TSDCT (18.57 dB, 20%) (b) TV (25.06 dB, 20%) (c) RCoS (31.33 dB, 20%)

(d) TSDCT (20.33 dB, 30%) (e) TV (28.95 dB, 30%) (f) RCoS (35.42 dB, 30%)

Figure 5: CS recovered portion of Image *Leaves* (20% and 30% measurements).

[2] E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inform. Theory*, vol. 52, pp. 489–509, 2006.

[3] E. Candès and M. Wakin, "An introduction to compressive sampling," *IEEE Signal Process. Magazine*, vol. 25, no. 2, pp. 21–30, 2008.

[4] L. He and L. Carin, "Exploiting structure in wavelet-based Bayesian compressive sensing," *IEEE Trans. Signal Process.*, vol. 57, no. 9, pp. 3488–3497, 2009.

[5] X. Wu, X. Zhang and J. Wang, "Model-guided adaptive recovery of compressive sensing," *Proc. of IEEE Data Compression Conf.*, pp. 123–132, 2009.

[6] S. Mun and J. E. Fowler, "Block compressed sensing of images using directional transforms," *Proc. of IEEE Int. Conf. Image Process.*, pp. 3021–3024, 2009.

[7] L. He, H. Chen and L. Carin, "Tree-structured compressive sensing with variational Bayesian analysis," *IEEE Signal Processing Letter*, vol. 17, no. 3, pp. 233–236, 2010.

[8] TVAL3 Toolbox available at http://www.caam.rice.edu/~optimization/L1/TVAL3/.

[9] Y. Kim, M. S. Nadar and A. Bilgin, "Compressed sensing using a Gaussian scale mixtures model in wavelet domain," *Proc. of IEEE Int. Conf. Image Process.*, pp. 3365–3368, 2010.

[10] D. Krishnan and R. Fergus, "Fast image deconvolution using hyper-Laplacian priors," *Proc. of Advancesin Neural Information Processing Systems*, vol. 22, pp. 1–9, 2009.

[11] S. Roth and M. J. Black, "Fields of experts," *International Journal of Computer Vision*, vol. 82, no. 2, pp. 205–229, 2009.

[12] A. Buades, B. Coll, and J. M. Morel, "A non-local algorithm for image denoising," *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 60–65, 2005.

[13] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3D transform-domain collaborative filtering," *IEEE Trans. on Image Process.*, vol. 16, no. 8, pp. 2080–2095, 2007.

[14] M. Afonso, J. Bioucas-Dias and M. Figueiredo, "Fast image recovery using variable splitting and constrained optimization," *IEEE Trans. on Image Process.*, vol. 19, no. 9, pp. 2345–2356, 2010.

[15] A. M. Bruckstein, D.L. Donoho, and M. Elad, "From sparse solutions of systems of equations to sparse modeling of signals and images," *SIAM Review*, vol. 51, no. 1, pp. 34–81, 2009.

[16] YALL1 Toolbox available at http://yall1.blogs.rice.edu/.

[17] l1-Magic Toolbox available at http://users.ece.gatech.edu/~justin/l1magic/.