# Compressed Sensing Dynamic Cardiac Cine MRI Using Learned Spatiotemporal Dictionary

Yanhua Wang and Leslie Ying*, *Senior Member, IEEE*

***Abstract*—In dynamic cardiac cine magnetic resonance imaging, the spatiotemporal resolution is limited by the low imaging speed. Compressed sensing (CS) theory has been applied to improve the imaging speed and thus the spatiotemporal resolution. In this paper, we propose a novel technique that employs a patch-based 3-D spatiotemporal dictionary for sparse representations of dynamic image sequence in the CS framework. Specifically, the dynamic image sequence is divided into overlapping patches along both the spatial and temporal directions. The dictionary is used to provide flexible sparse expressions for these patches. The underlying optimization problem is solved by variable splitting and the alternating direction method with multiplier. Experimental results based on *in vivo* cardiac data demonstrate that the proposed method is able to accelerate cardiac cine imaging by a factor up to 8 and outperforms the existing state-of-the-art CS methods at high accelerations. The method is expected to be useful in dynamic imaging with a higher spatiotemporal resolution.**

***Index Terms*—Alternating direction method with multiplier (ADMM), dynamic cardiac cine magnetic resonance imaging (MRI), spatiotemporal dictionary.**

## I. Introduction

DYNAMIC cardiac cine magnetic resonance imaging (MRI) is a powerful noninvasive and nonionizing technique to monitor the characteristics of cardiac motion [1]. Due to the physical and physiological constraints, there is a tradeoff between spatial and temporal resolutions in dynamic MRI (dMRI). To address the tradeoff, a variety of techniques have been devised, such as fast-scan techniques with advanced pulse sequences (e.g., EPI [2], [3], Spiral [4], [5]), parallel imaging using phased array coils [6]–[9], and reduced-scan techniques with advanced image reconstruction algorithms (e.g., partial Fourier [10], [11], RIGR [12], UNFOLD [13], view-sharing [14], partial separable function [15]–[17], and $k-t$ BLAST [18]). In reduced-scan techniques, images need to be reconstructed from undersampled $k$-space data, which leads to an ill-posed inverse problem. To overcome the ill-posedness, prior information is usually introduced in terms of certain mathematical models. Recently compressed sensing (CS) has emerged as a new theory to incorporate prior information in solving ill-posed problems and has demonstrated great potential for recovering a signal from its insufficient measurements [19], [20]. The CS theory imposes the sparseness as the prior constraint which asserts that the signal has only few nonzero elements in some transform domains. The signal can be thereby accurately recovered from its reduced measurements by some nonlinear reconstruction algorithms when the measurements satisfy certain conditions. As an emerging reduced-scan technique, CS has been successfully applied to accelerate dMRI acquisition [21]–[30]. Such success is made possible by two properties of dynamic MR images. First, strong correlations that exhibit between the images in the dynamic sequence enable sparse representations in specific transform domains. Second, the sampling scheme in *k-t* space can be elaborately designed to satisfy the conditions required by CS.

The sparsifying transform is essential for the success of CS reconstruction. The sparser the signal is made with a transform, the better the CS reconstruction is obtained. In the context of dMRI, a sparsifying transform can be applied along the spatial direction, the temporal direction or both. Typical spatial sparsifying transforms include finite difference and wavelet transform [31], and temporal transforms include Fourier transform [22], [23], [28]–[30], finite difference [26], [27], and principal component analysis (PCA) [16], [22], [32]. For the spatiotemporal transforms, 3-D transforms such as wavelet-Fourier transform [21] or 3-D wavelet transform [24] have been exploited. The abovementioned models fall into the category of predefined transforms. Such transforms have the advantage of simplicity with known properties, but might not be able to adequately capture all features of the specific image sequence of interest.

To this end, the studies of adaptive sparse representations have attracted significant research interests. One of the famous approaches is dictionary learning (DL) [33]–[35]. A properly trained overcomplete dictionary is able to provide a sparser representation over the fixed analytical transforms and has been used in various applications, such as image/video restoration [36], [37], machine learning [38], [39], and computer vision [40], [41]. In the context of CS for MRI, earlier approaches employed a dictionary to sparsify either the 2-D image patches [42]–[44] or the 1-D profiles of a sequence of MR images [45].

In this paper, we study a novel patch-based 3-D spatiotemporal dictionary for CS dMRI reconstruction. The dictionary is learned iteratively from 3-D overlapping patches of the dynamic image sequence along both the spatial and temporal directions.

Manuscript received August 9, 2013; revised October 29, 2013; accepted December 4, 2013. Date of publication December 12, 2013; date of current version March 17, 2014. This work was supported in part by the National Science Foundation CBET-0846514. *Asterisk indicates corresponding author.*

Y. Wang is with the Department of Biomedical Engineering and the Department of Electrical Engineering, University at Buffalo, The State University of New York, Buffalo, NY 14260 USA (e-mail: wyhlucky@gmail.com).

*L. Ying is with the Department of Biomedical Engineering and the Department of Electrical Engineering, University at Buffalo, The State University of New York, Buffalo, NY 14260 USA (e-mail: leiying@buffalo.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TBME.2013.2294939

Some preliminary results of this paper were partially presented in [46], [47]. Similar ideas were also presented independently in [48]–[51]. Our preliminary study has shown encouraging improvements of the 3-D DL approach over the conventional CS approaches. In this paper, we aim to provide a comprehensive study of the proposed method for dynamic cardiac image reconstruction using the 3-D DL model. We study the effect of additional total variation (TV) regularization on the image quality and how to choose the 3-D overlapping patch sizes. We also develop an efficient algorithm based on the alternating direction method with multiplier (ADMM) optimization scheme. Experimental results confirm that the proposed method with 3-D DL and TV regularization achieves superior reconstruction quality in comparison to the existing CS dMRI schemes.

This paper is structured as follows. Section II presents the signal model and reviews some related works on CS and DL. In Section III, we introduce the proposed patch-based 3-D DL reconstruction framework. The ADMM-based optimization algorithm is also detailed. In Section IV, the performance of the proposed method is assessed using five cardiac cine datasets. The robustness to noise and parameter selections is also investigated. Section V discusses the relationship to other existing works. Finally, the paper is concluded in Section VI.

## II. Background and Related Work

### A. Signal Model

In MR dynamic cardiac imaging, the measured signal at time point $t$ can be modeled as follows:

$$y(\mathbf{k},t) = \int_{\mathbf{r}} x(\mathbf{r},t) \exp(-j2\pi \mathbf{k}\cdot\mathbf{r})\, d\mathbf{r} + n(\mathbf{k},t) \quad (1)$$

where $\mathbf{r}$ and $\mathbf{k}$ are the locations in the spatial and $k$-space domain, respectively, $x(\mathbf{r},t)$ denotes the desired dynamic image series, and $n(\mathbf{k},t)$ represents the measured noise. In order to discretize (1), $x(\mathbf{r},t)$ and $y(\mathbf{k},t)$ are sampled on the grids of $N$ spatial locations $\{\mathbf{r}_i\}_{i=1}^{N}$ and $M$ $k$-space locations $\{\mathbf{k}_i\}_{i=1}^{M}$ ($M \leq N$). Then, we have the discrete form of (1):

$$\mathbf{y}_t = \mathbf{F}_t \mathbf{x}_t + \mathbf{n}_t \quad (2)$$

where $\mathbf{x}_t = [x(\mathbf{r}_1,t), x(\mathbf{r}_2,t),\ldots, x(\mathbf{r}_N,t)]^T$, $\mathbf{y}_t = [y(\mathbf{k}_1,t), y(\mathbf{k}_2,t), \ldots, y(\mathbf{k}_M,t)]^T$, $\mathbf{F}_t$ is the measurement matrix and $\mathbf{n}_t = [n(\mathbf{k}_1,t), n(\mathbf{k}_2,t), \ldots, n(\mathbf{k}_M,t)]^T$. More specifically, $\mathbf{F}_t = \mathbf{P}_t\mathbf{F}_{2-D}$, where $\mathbf{F}_{2-D}$ is an $N \times N$ unitary matrix representing the 2-D Fourier transform and $\mathbf{P}_t$ is an $M \times N$ matrix whose rows are extracted from an $N \times N$ identity matrix according to the $k$-space sampling locations at time $t$. Suppose a total of $Q$ cardiac phases are acquired. The entire data acquisition process can be formulated as follows:

$$\mathbf{y} = \mathbf{F}_u \mathbf{x} + \mathbf{n} \quad (3)$$

where $\mathbf{x} = [\mathbf{x}_1^T, \mathbf{x}_2^T,\ldots, \mathbf{x}_Q^T]^T$, $\mathbf{y} = [\mathbf{y}_1^T, \mathbf{y}_2^T,\ldots, \mathbf{y}_Q^T]^T$, $\mathbf{n} = [\mathbf{n}_1^T, \mathbf{n}_2^T,\ldots, \mathbf{n}_Q^T]^T$, and $\mathbf{F}_u = \mathrm{diag}\{\mathbf{F}_1, \mathbf{F}_2, \ldots, \mathbf{F}_Q\}$.

### B. Compressed Sensing Dynamic MRI

In accelerated imaging, the $k$-space data are undersampled ($M < N$) and therefore (3) is underdetermined. Regularizations are usually introduced to overcome the ill-posedness. The CS theory has provided a promising approach to address the issue of ill-posedness by imposing the sparsity regularizations [19], [20]. In the noiseless case, CS dMRI problem is formulated as the following $l_0$ minimization problem:

$$\min_{\mathbf{x}} \|\mathbf{\Psi x}\|_0 \quad \text{s.t.} \quad \mathbf{F}_u\mathbf{x} = \mathbf{y} \quad (4)$$

where $\|\mathbf{x}\|_0$ counts the number of nonzero elements of the vector and $\mathbf{\Psi}$ represents a certain sparsifying transform. The $l_0$ term enforces the sparsity in the transform domain and the equality constraint ensures the consistency with the measured data. The performance of CS reconstruction is guaranteed when $\mathbf{\Psi}$ and $\mathbf{F}_u$ satisfy certain conditions, such as the restricted isometric property [52], [53] and the mutual coherence property [54], [55].

The $l_0$ minimization problem in (4) is NP hard. A wide variety of approaches have been developed to mitigate the high computational complexity [56]. One major category is greedy pursuit algorithms, such as orthogonal matching pursuit (OMP) [57] and iterative hard thresholding [58]. Another important approach is to replace $l_0$ with $l_1$ [19], [20], $l_p$ ($0 < p < 1$) [59], or a hybrid quasi-norm [60]. Since $l_1$ minimization is convex and has many efficient solvers, it is widely employed in practical applications. The corresponding formulation is as follows:

$$\min_{\mathbf{x}} \|\mathbf{\Psi x}\|_1 \quad \text{s.t.} \mathbf{F}_u\mathbf{x} = \mathbf{y} \quad (5)$$

where $\|\mathbf{x}\|_1 = \Sigma|x_i|$. The exact recovery of (5) is guaranteed, but at the cost of increased measurements when compared to the original problem of (4) [52]. In the presence of Gaussian measurement noise, (5) is modified to:

$$\min_{\mathbf{x}} \|\mathbf{\Psi x}\|_1 \quad \text{s.t.} \|\mathbf{F}_u\mathbf{x} - \mathbf{y}\|_2^2 \leq M\sigma^2 \quad (6)$$

where $\sigma^2$ is the noise power. Rewriting the formulation using Lagrangian, we obtain the unconstrained form:

$$\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{F}_u\mathbf{x} - \mathbf{y}\|_2^2 + \lambda \|\mathbf{\Psi x}\|_1 \quad (7)$$

where $\lambda$ is the penalty parameter related to the noise power.

The sparsifying transform $\mathbf{\Psi}$ is crucial to the reconstruction process. dMRI data exhibits redundancies in both spatial and temporal directions. Thus, $\mathbf{\Psi}$ can be selected to enforce the spatial sparsity, temporal sparsity, or spatiotemporal sparsity. The most straightforward scheme simply applies the CS MRI techniques frame by frame [31] where $\mathbf{\Psi}$ is only used to sparsify the images spatially. On the other hand, the dynamic image sequence can be sparsified in time because it exhibits high correlation along the temporal direction. Typically used transforms in temporal direction include Fourier transform (e.g., [22], [23], [28]–[30]), finite difference (e.g., [27]), and PCA (e.g., [16], [22], [32]). The spatial and temporal constraints can be combined together to exploit the spatial and temporal correlations simultaneously [26]. Moreover, the spatiotemporal sparsifying transforms are also appealing to exploit such correlations. Typical transforms include wavelet-Fourier transform [21] and 3-D wavelet transform [24].

### C. Dictionary Learning in Compressed Sensing Dynamic MRI

Besides the predefined sparsifying transforms which are applicable to specific signals, many researchers have investigated the dictionary-based sparse representation model [33]–[35]. A dictionary is a collection of "basic" signals used to sparsely represent a given class of signals. Each "basic" signal is called an atom. In most cases, the dictionary is overcomplete (or redundant) which means that the number of atoms is larger than the dimension of a single atom. Representing a signal using a dictionary can take either the synthesis way or the analysis form. Extensive research studies focus on the synthesis framework where a given signal is expressed as the linear combination of a few atoms [33], [34]. In contrast, the analysis approach treats the dictionary as a linear operator [34], [35]. The sparse coefficients of a given signal are computed via its inner products with the atoms. The synthesis and analysis models are dramatically different and we will focus on the synthesis path in the following discussion.

A data adaptive dictionary is required for specific datasets. The process of training a dictionary from datasets with the same features is called DL [33], [34]. Mathematically, given a set of training signals $\{\mathbf{s}_j\}_{j=1}^{J}$, DL aims to solve the following problem:

$$\min_{\mathbf{D},\{\boldsymbol{\alpha}_j\}} \sum_{j=1}^{J} \|\mathbf{s}_j - \mathbf{D}\boldsymbol{\alpha}_j\|_2^2 \quad \text{s.t.}\, \|\boldsymbol{\alpha}_j\|_0 \leq K\, \forall j \tag{8}$$

where $\mathbf{D}$ stands for the dictionary to be constructed, $\boldsymbol{\alpha}_j$ is called the sparse representation of $\mathbf{s}_j$ over $\mathbf{D}$, and $K$ represents the sparsity constraint. As an overcomplete dictionary, $\mathbf{D}$ has more columns than rows. Each atom in $\mathbf{D}$ is set to be a unit-norm vector in order to handle the scaling ambiguity. The above constrained optimization problem can be rewritten in the following unconstrained form [36], [37]:

$$\min_{\mathbf{D},\{\boldsymbol{\alpha}_j\}} \frac{1}{2} \sum_{j=1}^{J} \|\mathbf{s}_j - \mathbf{D}\boldsymbol{\alpha}_j\|_2^2 + \mu \sum_{j=1}^{J} \|\boldsymbol{\alpha}_j\|_0. \tag{9}$$

The two formulations in (8) and (9) are equivalent when the penalty parameter $\mu$ is properly selected. The DL problem is nonconvex and NP hard. Typical algorithms adopt the alternating optimization approach which iteratively updates $\mathbf{D}$ and $\boldsymbol{\alpha}_j$ [61], [62]. While the $\boldsymbol{\alpha}$-step can be solved by various sparse coding algorithms, the core difference lies in the $\mathbf{D}$-step. A survey of typical DL algorithms can be found in [33] and [34].

Dictionary-based sparse representation or DL has recently been applied for CS dMRI reconstructions. Since the computational complexity of DL algorithms is usually very high, practical reconstruction models often enforce the "patch-level" sparsity. A data patch can be extracted along the spatial direction or the temporal direction.

When only the spatial correlation is exploited, the dynamic images are reconstructed frame by frame [42]–[44]. In [42], the dictionary is trained from prescanned images and the desired images are reconstructed by solving the following problem:

$$\min_{\mathbf{x},\{\boldsymbol{\alpha}_j\},\sigma} \mathrm{TV}_S(\mathbf{x}) + \mu \sum_j \left( \|\boldsymbol{\alpha}_j\|_1 + \frac{\nu}{2} \|\mathbf{R}_j\mathbf{x} - \mathbf{D}\boldsymbol{\alpha}_j\|_2^2 \right) + \lambda F(\mathbf{x},\sigma) \tag{10}$$

where $\mathrm{TV}_S(\mathbf{x})$ is the isotropic spatial TV, $\mathbf{R}_j$ extracts a square image patch at a certain time point, $\mathbf{F}(\mathbf{x},\sigma) = \|\mathbf{F}_u\mathbf{x} - \mathbf{y}\|_2^2/2\sigma^2 + M\times\log\sigma$. In contrast, the dictionary is adaptively learned from the measured data in [43] and the formulation becomes:

$$\min_{\mathbf{x},\mathbf{D},\{\boldsymbol{\alpha}_j\}} \sum_j \|\mathbf{R}_j\mathbf{x} - \mathbf{D}\boldsymbol{\alpha}_j\|_2^2 + \nu\|\mathbf{F}_u\mathbf{x} - \mathbf{y}\|_2^2 \ \text{s.t.} \|\boldsymbol{\alpha}_j\|_0 \leq K\, \forall j. \tag{11}$$

The alternating optimization method is applied to sequentially update $\mathbf{D}$, $\boldsymbol{\alpha}_j$, and $\mathbf{x}$. An alternate form of (11) is proposed in [44]:

$$\min_{\mathbf{x},\mathbf{D},\{\boldsymbol{\alpha}_j\}} \sum_j \left( \|\boldsymbol{\alpha}_j\|_1 + \frac{\nu}{2} \|\mathbf{R}_j\mathbf{x} - \mathbf{D}\boldsymbol{\alpha}_j\|_2^2 \right) \quad \text{s.t.} \quad \|\mathbf{F}_u\mathbf{x} - \mathbf{y}\|_2^2 \leq M\sigma^2 \tag{12}$$

and a Bregman iterative algorithm is developed to reduce the computational load and accelerate the convergence.

The dictionary has also been used to represent the temporal profiles of a dynamic image sequence. In [45], DL is applied for CS reconstruction of parameter mapping where the temporal curves are modeled by the exponential function. The dictionary is learned from a large collection of prototype signals generated from the model. The image sequence and its sparse coefficients are then iteratively updated based on the fixed dictionary.

## III. Proposed Method

### A. Problem Formulation

In this paper, we investigate the CS dynamic cardiac MRI using 3-D DL model for sparse representation. Let $\mathbf{R}_j$ be the operator that extracts a spatiotemporal patch from the dynamic image sequence $\mathbf{x}$ with a predefined size of $(n_f, n_p, n_t)$. The 3-D DL problem is expressed as follows:

$$\min_{\mathbf{D},\{\boldsymbol{\alpha}_j\}} \frac{1}{2} \sum_j \|\mathbf{R}_j\mathbf{x} - \mathbf{D}\boldsymbol{\alpha}_j\|_2^2 + \mu \sum_j \|\boldsymbol{\alpha}_j\|_0. \tag{13}$$

We use the unconstrained form to simplify the expressions in the following derivations. The first term captures the representation errors and the second term implies the sparsity level. The dictionary $\mathbf{D}$ has a size of $d_r \times d_c$, where $d_r = n_f \times n_p \times n_t$ and $d_c$ is the number of atoms. The overcompleteness (or redundancy) $\eta$ is defined as $\eta = d_c/d_r$. The variation of index $j$ covers all the available patches. The patch stride is measured by a vector $\mathbf{p} = [p_f, p_p, p_t]^T$ which indicates the distance between the starting voxels in two adjacent patches. In this paper, we use periodically positioned overlapping patches. We adopt the maximum overlap ($\mathbf{p} = [1, 1, 1]^T$), and abide by the periodic boundary conditions. The dimensions of $\mathbf{x}$ are extended before a patch goes beyond the boundaries. The extension parts are filled by the data from the opposite sides of the image sequence.

Based on the aforementioned 3-D DL model, we formulate the CS dMRI problem as follows:

$$\min_{\mathbf{x},\mathbf{D},\{\boldsymbol{\alpha}_j\}} \frac{1}{2}\left\|\mathbf{F}_u\mathbf{x}-\mathbf{y}\right\|_2^2+\frac{\lambda_1}{2}\sum_j\left\|\mathbf{R}_j\mathbf{x}-\mathbf{D}\boldsymbol{\alpha}_j\right\|_2^2 + \mu\sum_j\left\|\boldsymbol{\alpha}_j\right\|_0+\lambda_2 h\left(\mathbf{x}\right) \tag{14}$$

where $\lambda_1$ and $\lambda_2$ are tuning parameters, and $h(\mathbf{x})$ is an additional regularizer. The first term enforces the data fidelity in $k$-space. The second and the third terms correspond to the 3-D DL model. The selection of $h(\mathbf{x})$ can be any suitable sparsifying transforms, such as TV or temporal Fourier transform. In this paper, we choose the general anisotropic 3-D TV [63]:

$$h\left(\mathbf{x}\right)=\left\|\mathbf{G}\mathbf{x}\right\|_1=\beta_x\left\|\mathbf{G}_x\mathbf{x}\right\|_1+\beta_y\left\|\mathbf{G}_y\mathbf{x}\right\|_1+\beta_t\left\|\mathbf{G}_t\mathbf{x}\right\|_1 \tag{15}$$

where $\mathbf{G}_x$, $\mathbf{G}_y$, and $\mathbf{G}_t$ are first-order forward finite difference operators along the frequency encoding (FE), phase encoding (PE), and temporal directions, respectively, $\mathbf{G}$ is the collection of $\mathbf{G}_x$, $\mathbf{G}_y$, and $\mathbf{G}_t$, i.e., $\mathbf{G}=[\beta_x\mathbf{G}_x^T,\beta_y\mathbf{G}_y^T,\beta_z\mathbf{G}_z^T]^T$, $\beta_x$, $\beta_y$, and $\beta_t$ are constant scalars.

### B. Optimization Algorithm

It is challenging to simultaneously find all unknown variables in the objective function (14) which is not differentiable due to the $l_0$ and $l_1$ term. It is also nonconvex because of the $l_0$ term and the multiplication between $\mathbf{D}$ and $\boldsymbol{\alpha}_j$. We note that the nonconvex part (optimization with respect to $\mathbf{D}$ and $\boldsymbol{\alpha}_j$) is independent of other terms once $\mathbf{x}$ is fixed. This part can be solved by any DL algorithms. When $\mathbf{D}$ and $\boldsymbol{\alpha}_j$ are known, (14) is simplified to a standard $l_1$ minimization problem. Recently, variable splitting and augmented Lagrangian method (ALM) (or the Bregman iteration method) have proven to be efficient approaches to solve the $l_1$ minimization problems [64]–[66]. In general, these algorithms introduce auxiliary variables to decouple the sparsity terms with other parts. ALM is then applied to solve the modified constrained problem. In this paper, we focus on a specific variation of ALM called ADMM [65].

We first introduce an auxiliary variable $\mathbf{d}=\mathbf{G}\mathbf{x}$ to decouple the $l_1$ term from other parts and obtain the following equivalent problem:

$$\min_{\mathbf{d},\mathbf{x},\mathbf{D},\{\boldsymbol{\alpha}_j\}} \frac{1}{2}\left\|\mathbf{F}_u\mathbf{x}-\mathbf{y}\right\|_2^2+\frac{\lambda_1}{2}\sum_j\left\|\mathbf{R}_j\mathbf{x}-\mathbf{D}\boldsymbol{\alpha}_j\right\|_2^2 + \mu\sum_j\left\|\boldsymbol{\alpha}_j\right\|_0+\lambda_2\left\|\mathbf{d}\right\|_1 \quad \text{s.t. } \mathbf{d}=\mathbf{G}\mathbf{x}. \tag{16}$$

The scaled augmented Lagrangian function of (16) is [65]:

$$L_\rho\left(\mathbf{d},\mathbf{x},\mathbf{D},\boldsymbol{\alpha}_j,\mathbf{u}\right)=\frac{1}{2}\left\|\mathbf{F}_u\mathbf{x}-\mathbf{y}\right\|_2^2+\frac{\lambda_1}{2}\sum_j\left\|\mathbf{R}_j\mathbf{x}-\mathbf{D}\boldsymbol{\alpha}_j\right\|_2^2 +\mu\sum_j\left\|\boldsymbol{\alpha}_j\right\|_0+\lambda_2\left\|\mathbf{d}\right\|_1+\frac{\rho}{2}\left\|\mathbf{d}-\mathbf{G}\mathbf{x}+\mathbf{u}\right\|_2^2 \tag{17}$$

where $\mathbf{u}$ is called the scaled dual variable of the Lagrangian multiplier and $\rho$ is the penalty parameter. Based on ADMM, we derive the following iteration scheme:

$$\left\{\mathbf{D}^{k+1},\boldsymbol{\alpha}_j^{k+1}\right\}=\arg\min_{\mathbf{D},\{\boldsymbol{\alpha}_j\}}\frac{\lambda_1}{2}\sum_j\left\|\mathbf{R}_j\mathbf{x}^k-\mathbf{D}\boldsymbol{\alpha}_j\right\|_2^2 +\mu\sum_j\left\|\boldsymbol{\alpha}_j\right\|_0 \tag{18}$$

$$\mathbf{x}^{k+1}=\arg\min_{\mathbf{x}}\frac{1}{2}\left\|\mathbf{F}_u\mathbf{x}-\mathbf{y}\right\|_2^2+\frac{\lambda_1}{2}\sum_j\left\|\mathbf{R}_j\mathbf{x}-\mathbf{D}^{k+1}\boldsymbol{\alpha}_j^{k+1}\right\|_2^2 +\frac{\rho}{2}\left\|\mathbf{d}^k-\mathbf{G}\mathbf{x}+\mathbf{u}^k\right\|_2^2 \tag{19}$$

$$\mathbf{d}^{k+1}=\arg\min_{\mathbf{d}}\lambda_2\left\|\mathbf{d}\right\|_1+\frac{\rho}{2}\left\|\mathbf{d}-\mathbf{G}\mathbf{x}^{k+1}+\mathbf{u}^k\right\|_2^2 \tag{20}$$

$$\mathbf{u}^{k+1}=\mathbf{u}^k+\mathbf{d}^{k+1}-\mathbf{G}\mathbf{x}^{k+1}. \tag{21}$$

We now delineate the way of solving the subproblems (18)–(20).

*1) $\{\mathbf{D},\boldsymbol{\alpha}\}$-Subproblem:* The minimization in (18) is the unconstrained form of DL problem. We first convert it to the constrained form [36], [37]:

$$\left\{\mathbf{D}^{k+1},\boldsymbol{\alpha}_j^{k+1}\right\}=\arg\min_{\mathbf{D},\{\boldsymbol{\alpha}_j\}}\sum_j\left\|\mathbf{R}_j\mathbf{x}^k-\mathbf{D}\boldsymbol{\alpha}_j\right\|_2^2 \quad \text{s.t. } \left\|\boldsymbol{\alpha}_j\right\|_0\le K\,\forall j. \tag{22}$$

As noted in the previous section, the two forms are equivalent for proper values of $\mu$ and $K$. This way, the choice of $\mu$ in (18) (also in (14)) is handled implicitly when we determine $K$ [36], [37]. Choice of $K$ is studied in Section IV. The constrained DL problem (22) can be efficiently solved by the K-SVD algorithm [62]. It adopts a two-stage optimization process which performs several rounds of alternating updates between $\boldsymbol{\alpha}_j$ and $\mathbf{D}$. The $\boldsymbol{\alpha}_j$ stage is a sparse coding problem which is solved by the OMP algorithm. Then, $\mathbf{D}$ is updated column by column through a rank-one approximation of a residual matrix by singular value decomposition (SVD). The K-SVD algorithm needs an initial guess of $\mathbf{D}$. We use the overcomplete discrete cosine transform (DCT) dictionary for initialization. Although the K-SVD algorithm cannot guarantee to reach a global minimum, it still shows excellent performances in many applications [36], [37]. Since the amount of overlapping patches in $\mathbf{x}$ is usually very large, we first randomly select a small portion of all patches for dictionary training. Once $\mathbf{D}$ is learned from the training dataset, we apply OMP to compute $\boldsymbol{\alpha}_j$ for all patches.

*2) $\mathbf{x}$-Subproblem:* This least-squares problem is equivalent to the following linear equation:

$$\left(\mathbf{F}_u^H\mathbf{F}_u+\lambda_1\sum_j\mathbf{R}_j^T\mathbf{R}_j+\rho\mathbf{G}^T\mathbf{G}\right)\mathbf{x} =\mathbf{F}_u^H\mathbf{y}+\lambda_1\sum_j\mathbf{R}_j^T\mathbf{D}\boldsymbol{\alpha}_j+\rho\mathbf{G}^T\left(\mathbf{d}+\mathbf{u}\right). \tag{23}$$

Here, we drop iteration indices to simplify the expression. The superscripts $H$ and $T$ denote the Hermitian transpose and the transpose for a matrix, respectively. The operator $\mathbf{F}_u^H$ fills the missing entries with zero and performs the inverse Fourier

transform. Hence, $\mathbf{F}_u^H\mathbf{y}$ is the zero filling reconstruction of the measured data. The operator $\mathbf{R}_j^T$ places a patch back to its original position and $\sum_j \mathbf{R}_j^T\mathbf{D}\boldsymbol{\alpha}_j$ is the assembly of all estimated patches. The term $\sum_j \mathbf{R}_j^T\mathbf{R}_j$ is a diagonal matrix where the value of each diagonal entry is equal to the number of the patches that contain the corresponding voxel. Since we apply the maximum overlapping patches and the periodic boundary conditions, the diagonal entries are all equal to $d_r$, i.e., $\sum_j \mathbf{R}_j^T\mathbf{R}_j = d_r\mathbf{I}$.

Equation (23) has a closed-form solution by left-multiplying the inverse of the coefficient matrix on both sides, but the computation of the inverse matrix is impractical due to its large size. Therefore, we apply the preconditioned conjugated gradient (PCG) algorithm to solve this problem. In order to improve the convergence, the PCG is initialized with the result from the previous iteration.

*3) **d**-Subproblem:* In this problem, there is no coupling between elements of $\mathbf{d}$. The optimal value is obtained through the element-wise shrinkage operator [64]:

$$d_i^{k+1} = \text{shrink}\left(\left[\mathbf{G}\mathbf{x}^k\right]_i - u_i^k, \frac{\lambda_2}{\rho}\right) \tag{24}$$

where

$$\text{shrink}\,(x,\gamma) = \frac{x}{|x|}\max\left(|x| - \gamma, 0\right). \tag{25}$$

Although the aforementioned discussion is focused on 3-D TV, the algorithm can be easily modified to accommodate other forms of $l_1$-norm regurlarizer $h(\mathbf{x})$.

Algorithm 1 shows the pseudocode. The stopping criterion of the algorithm is either $||\mathbf{x}^{k+1} - \mathbf{x}^k||_2/||\mathbf{x}^k||_2 < \varepsilon$ or a predefined maximum iteration number is reached.

**Algorithm 1**

**Input:** $\mathbf{y}$ – undersampled k-space measurements
$\lambda_1$, $\lambda_2$, $\rho$ – regularization parameters
$\beta_x$, $\beta_y$, $\beta_t$ – TV weighting parameters
$n_f$, $n_p$, $n_t$ – patch sizes, $\eta$ – dictionary redundancy
$K$ – DL sparsity constraint
$\varepsilon$ – stopping threshold, $T$ – maximum iterations
**Output:** $\mathbf{x}$ – reconstructed dMRI image sequence
**Initialization:** $\mathbf{x} = \mathbf{F}_u^H\mathbf{y}$, $\mathbf{d} = \mathbf{0}$, $\mathbf{u} = \mathbf{0}$, $\mathbf{D}$ = overcomplete DCT dictionary
**While** $\|\mathbf{x}^{k+1} - \mathbf{x}^k\|_2/\|\mathbf{x}^k\|_2 < \varepsilon$ **and** $k < T$
1) Learn dictionary $\mathbf{D}$ from a subset of patches using K-SVD algorithm.
2) Compute $\boldsymbol{\alpha}_j$ for all patches using OMP.
3) Solve the $\mathbf{x}$-subproblem using PCG algorithm.
4) Solve the $\mathbf{d}$-subproblem using (24).
5) Update $\mathbf{u}$.
**End while**

## IV. Experimental Results

### A. Experimental Setup

The performances of the proposed method were evaluated using five sets of fully sampled 2-D dynamic cardiac cine MRI data. Prospective cardiac gating was used for all datasets during acquisition. Some sample frames reconstructed from original data are presented in Fig. 1 and the corresponding imaging parameters are listed in Table I. In order to simplify the parameter-tuning process, the five fully sampled datasets were first rescaled to the image intensity interval of [0, 1]. Then, a portion of the $k$-space data was manually extracted to simulate the undersampled scans. We define the net reduction $R$ as the ratio between the number of the fully sampled data points to that of the undersampled data points. Reconstructions on the same datasets using the *k-t* FOCUSS method [22], [28] were performed for comparison. The same variable-density random sampling pattern was used for both methods. For the datasets acquired using multiple coils, we reconstructed images for each coil separately and then combined them using square-root of sum-of-squares.

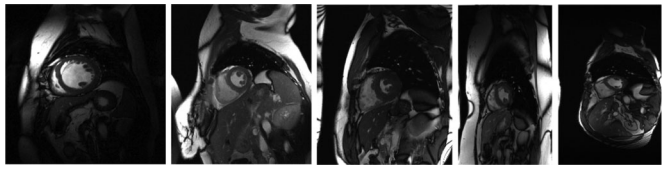

Fig. 1. Sample frames of fully sampled reconstructions. From left to right: datasets 1–5.

The $k$-space sampling pattern is critically important in CS reconstruction. The incoherence condition of the CS theory requires that the $k$-space data should be randomly sampled to generate incoherent artifacts in the sparse transform domain [31], [54], [55]. The ideal sampling scheme is the 2-D random pattern, but it is impractical [31]. Thus, we applied the 1-D undersampling pattern along the PE direction frame by frame [21], [22], [28]–[30]. More specifically, in each frame, the central part of the $k$-space was fully sampled with a total of eight PE lines. Other PE lines were randomly selected according to a given distribution. To achieve the randomness in *k-t* domain, the random patterns for different frames were generated independently with the same distribution. Here, we employed the Gaussian distribution which was used in the original *k-t* FOCUSS papers [22], [28]. Such a variable-density sampling pattern with more samples at the central $k$-space is preferred because most $k$-space energy is concentrated in the low-frequency part. Comparisons with other distributions such as uniform and polynomial distributions [29], [31] are studied in Section IV-D.

The optimal selection of the algorithm parameters is also crucial for the reconstruction process, but it is still an open problem. Optimal choices depend on the image structures, dynamic ranges, and motion characteristics. In this paper, we adopted an empirical approach. We first tuned the parameters (including the weight parameters in (14) and the DL parameters in (22)) for dataset 1 to reach a balance between reconstruction error and computational time. The same values were then applied for all other datasets. The detailed procedure and the effect of different parameters are discussed in the parameter evaluation section.

Throughout all simulations, the nominal values of the regularization parameters are $\lambda_1 = 10^{-2}$, $\lambda_2 = 10^{-4}$, and $\rho = 5\times10^{-3}$. As for the TV weights, we empirically adopt $\beta_x = \beta_y = 1$ and $\beta_t = 10$. The maximum iteration number $T$ is 25 and the

TABLE I
DATA ACQUISITION PARAMETERS FOR FIVE DATASETS

| Parameter | Dataset1 | Dataset2 | Dataset3 | Dataset4 | Dataset5 |
|---|---|---|---|---|---|
| Scanner | 1.5T Philip | 3T Siemens | 3T Siemens | 3T Siemens | 3T Siemens |
| Sequence | SSFP | SSFP | SSFP | SSFP | SSFP |
| Flip angle (°) | 50 | 50 | 50 | 50 | 44 |
| Repetition time / Echo time (ms) | 3.45/1.7 | 29.9/1.87 | 56.6/1.89 | 3.45/1.7 | 42.5/1.2 |
| Matrix size (PE × FE × Coil × Frame) | 256×256×1×25 | 224×256×4×14 | 216×256×15×19 | 150×256×4×17 | 165×304×12×26 |

stopping criteria $\varepsilon$ is $10^{-6}$. If not specifically mentioned, the default patch size is $n_f \times n_p \times n_t = 4 \times 4 \times 4$. In DL stage, we empirically set the dictionary overcompleteness factor $\eta$ to 4, i.e., $d_c = 4\, d_r$, and randomly select 50 $d_c$ patches as the training dataset. In the K-SVD algorithm, we employ ten iterations and the sparsity constraint $K$ is set to 15. We notice that these values of parameters work well for all datasets and the results are not very sensitive to slight changes in these values.

Reconstructions from the rescaled fully sampled data were used as the references. The performance was quantitatively evaluated in terms of the mean-squared error (MSE), which was calculated by

$$\mathrm{MSE}(\hat{\mathbf{x}}_t) = \|\hat{\mathbf{x}}_t - \mathbf{x}_t\|_2^2 \Big/ N \tag{26}$$

where $\hat{\mathbf{x}}_t$ is the reconstructed image at the $t$th frame. The MSE of the whole sequence was simply the average MSE of all frames. All programs were implemented in MATLAB R2012 a on a workstation with 3.4-GHz Intel Core i7 central processing unit, 16 GB of memory and Windows 7 Enterprise 64-bit operating system.

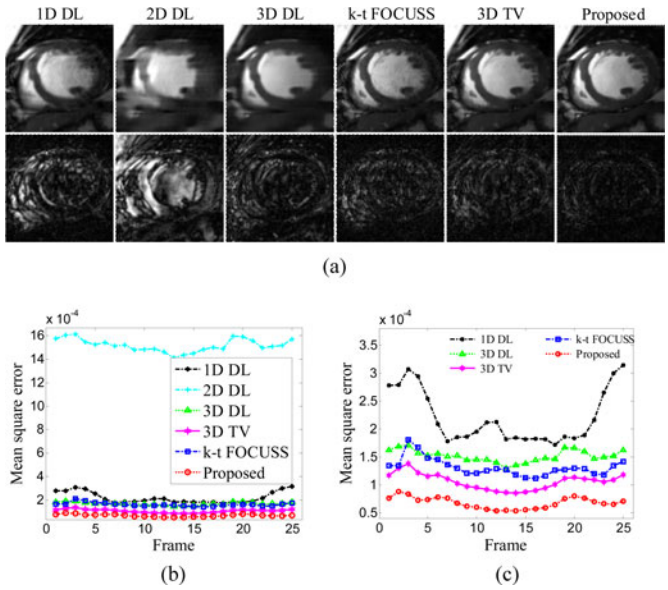


Fig. 2. Comparisons of reconstructions using different models (dataset 1, $R = 8$). (a) (Top row) ROI images and (bottom row) errors. (b) Comparison of MSE. (c) Zoom-in view of (b).

### B. Rationale Behind the Proposed Method

As stated in Section II, DL was previously introduced to CS dMRI to sparsify either the 2-D image patches (2-D DL) [42]–[44] or the 1-D signal profiles for parameter mapping (1-D DL) [45]. The two models only capture the spatial or temporal correlation alone in the image sequence. In contrast, the proposed 3-D DL model is a generalized formulation for dynamic imaging and it is expected to preserve the spatiotemporal correlation.

The sparse regularizers in the objective function (14) consist of two parts: one is the 3-D DL and the other is an additional sparsifying transform. Although we select 3-D TV as the additional term, it can be any other suitable operators. The parameters $\lambda_1$ and $\lambda_2$ control their contributions. When $\lambda_1 = 0$, (14) is boiled down to the conventional CS model. When $\lambda_2 = 0$, the objective function turns out to be the basic 3-D DL model.

In this experiment, we briefly compared the performances of 1-D DL, 2-D DL, and 3-D DL (proposed method without TV), 3-D TV, *k-t* FOCUSS, and the propose method. We reconstructed the images of dataset 1 using these models. The reduction factor was $R = 8$ and we fixed the iteration number to 40. Fig. 2(a) displays the region of interest (ROI) images and errors of the eleventh frame. The result of 1-D DL captures most of the myocardium contours but fails to suppress the artifacts. The image of 2-D DL is overly smoothed which indicates that it is unable to reconstruct data with such a high reduction factor. The 3-D DL model shows great superiority to 1-D DL and 2-D DL. The result is devoid of aliasing artifacts but loses some details due to the smoothing effect of the patch average operation. Both *k-t* FOCUSS and 3-D TV capture more details than 3-D DL but they still show some aliasing artifacts. The proposed method incorporates the advantages of 3-D DL and 3-D TV and thus produces the best reconstructions among all. Fig. 2(b) and (c) show the frame-by-frame MSEs of different models. As can be seen, the MSE of 2-D DL is the highest. The 1-D DL model achieves much lower MSE compared with 2-D DL, which implies the importance of exploiting temporal correlation. The results of 3-D DL, *k-t* FOCUSS and 3-D TV show MSE improvement to 1-D DL. The proposed method achieves the lowest MSE through the combination of 3-D DL and 3-D TV.

### C. Reconstruction Performance

In this experiment, we compared the reconstruction quality of the proposed method to that of *k-t* FOCUSS. We first reconstructed the image sequence of dataset 1 with $R = 8$. Fig. 3 shows the ROI images in three different frames (indices: 11, 16, 25) and the corresponding errors. The proposed method captures most of the structures, though some tiny details are smoothed due to the average of patches. Compared with *k-t* FOCUSS,

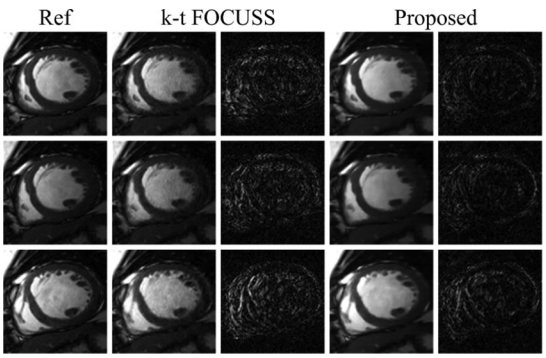


Fig. 3. ROI images and errors in different frames of dataset 1 ($R = 8$).

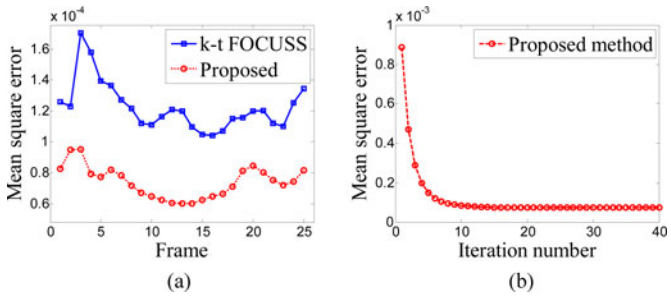


Fig. 4. Performance evaluation for dataset 1. (a) Comparison of MSE ($R = 8$). (b) MSE versus iteration number of proposed method.

the proposed method suppresses the PE direction artifacts and presents clear contours of the object. The error images with respect to the reference also demonstrate the superiority of the proposed method over *k-t* FOCUSS. While both methods show visible errors on the edges of structures, *k-t* FOCUSS leads to more aliasing errors. Moreover, the magnitudes of errors in *k-t* FOCUSS are higher than the proposed method. In order to provide a quantitative comparison on the reconstruction errors, Fig. 4(a) depicts the MSEs frame by frame. It is evident that the proposed method has lower MSEs than *k-t* FOCUSS for all frames.

We also evaluated the convergence of the proposed algorithm. The maximum iteration number of 40 was used as the stopping criterion in this test. We computed the MSE of the entire image sequence for each iteration. Fig. 4(b) plots the MSE curve versus the iteration number. The MSE decreases rapidly within the first few iterations and nearly reaches to a constant in the end. It can be seen that the MSE change between two successive iterations is very small after about 10 iterations. It suggests empirically that the proposed algorithm is able to converge fast.

Comparisons with *k-t* FOCUSS using other datasets were also performed for further validation. Fig. 5 shows the ROI images and errors of a middle frame in datasets 2–dataset 5. The net reduction factors were $R = 6$ for dataset 4 and $R = 8$ for other datasets. Compared with *k-t* FOCUSS, the proposed method shows much fewer artifacts and preserves sharper myocardial contours. Fig. 6 depicts the frame-by-frame MSEs. It is evident that the proposed method achieves lower MSEs than *k-t* FOCUSS for all datasets.

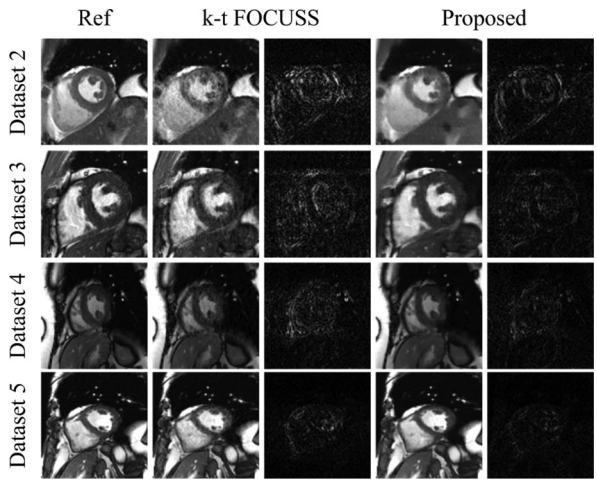


Fig. 5. ROI images and errors of different datasets.

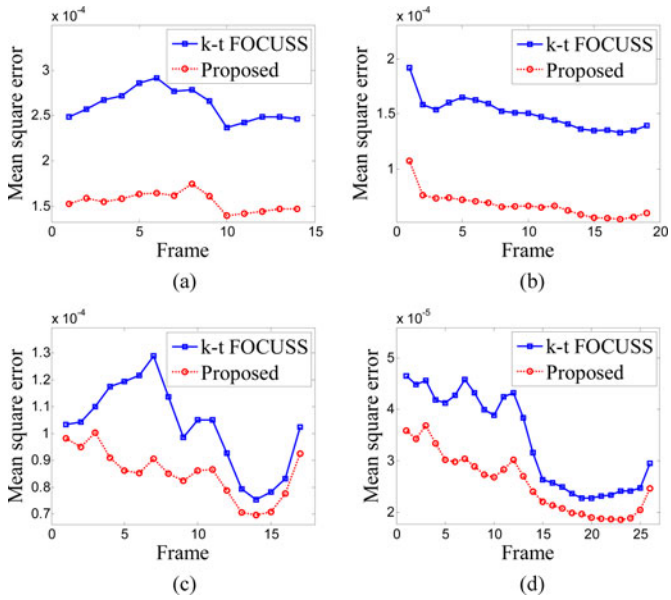


Fig. 6. Comparison of MSE for different datasets. (a) Dataset 2 ($R = 8$). (b) Dataset 3 ($R = 8$). (c) Dataset 4 ($R = 6$). (d) Dataset 5 ($R = 8$).

Fig. 7 compares the temporal profiles at a fixed pixel in the FE direction of datasets 2 and 3, both with $R = 8$. It is observed that the *k-t* FOCUSS reconstructions are blurry with some loss of variations and contrasts. The proposed method follows the original temporal variations closely and captures major motion features. In general, the proposed method achieves a greater level of image fidelity compared with *k-t* FOCUSS.

Table II compares the reconstruction time (only for a single channel) of *k-t* FOCUSS and the proposed method for all datasets. The iteration number was fixed to 25 for both methods. For each dataset, we performed ten reconstructions and computed the average time. For multicoil datasets, the total computational cost is the product of the single-channel cost and the number of coils. It is shown that the reconstruction time of the proposed method is about eight to nine times higher than that of *k-t* FOCUSS. The major computational burden comes from the great number of patches in the DL step and the large-scale linear

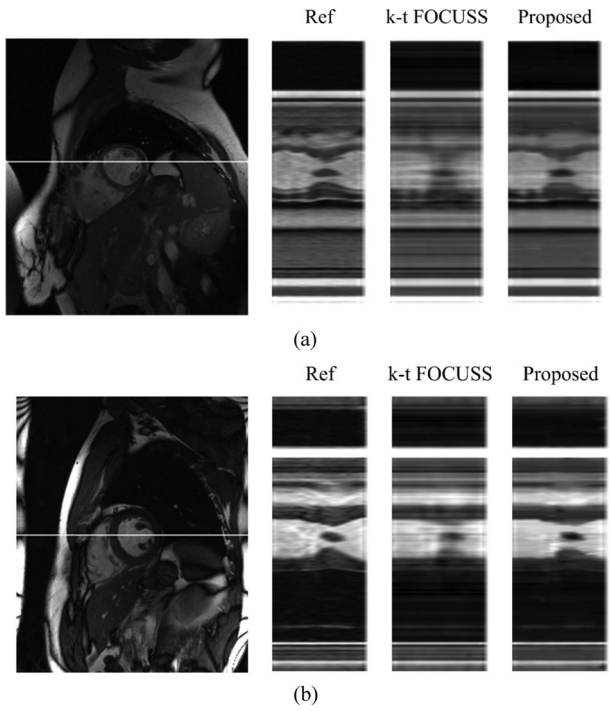


Fig. 7. Comparison of temporal profiles. (a) Dataset 2 ($R = 8$). (b) Dataset 3 ($R = 8$).

TABLE II
SINGLE-CHANNEL RECONSTRUCTION TIME OF FIVE DATASETS

| Dataset | k-t FOCUSS | Proposed |
|---|---|---|
| Dataset1 | 169 s | 1459 s |
| Dataset2 | 106 s | 897 s |
| Dataset3 | 114 s | 933 s |
| Dataset4 | 83 s | 680 s |
| Dataset5 | 156 s | 1423 s |

equation in the image update step. Since we adopt maximum patch overlap, the number of patches is very large, resulting in high computational time of $\{\boldsymbol{\alpha}_j\}$. We notice that the computations of $\{\boldsymbol{\alpha}_j\}$ for different patches are independent. Thus, the computational time can be reduced by parallel processing techniques. The computational load and the convergence rate of DL are also expected to be improved by the Bregman-based algorithm [44]. To handle the large scale of the linear equation in the $\mathbf{x}$ step, more auxiliary variables can be added so that the coefficient matrix is separated into several structured matrices with fast implementations of matrix inversion [67]. Such a strategy will accelerate the computation significantly, but at the expense of tuning more parameters.

### D. Sampling Patterns

We evaluated the reconstruction MSEs under the sampling patterns generated from three different distributions: uniform distribution, polynomial distribution [29], [31], and Gaussian distribution [22], [28]. Examples of sampling patterns in *k-t* domain are shown in Fig. 8(a). We reconstructed dataset 1 with $R = 8$ using *k-t* FOCUSS and the proposed method. For each

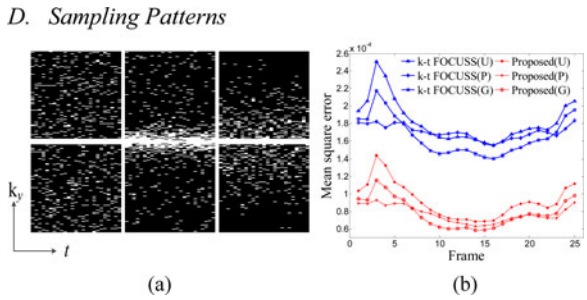


Fig. 8. Sampling pattern comparisons. (a) Different *k-t* patterns. Left: uniform distribution, middle: polynomial distribution, right: Gaussian distribution. (b) Comparison of MSE (U: uniform, P: polynomial, and G: Gaussian).

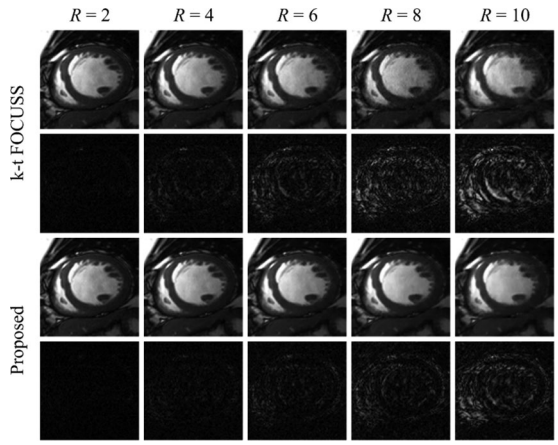


Fig. 9. ROI images and errors of the 11th frame in dataset 1 from $R = 2$ to $R = 10$.

distribution, we generated ten realizations of the random pattern, and the average frame-by-frame MSE curves are depicted in Fig. 8(b). Both methods are seen to favor Gaussian distribution which gives the lowest MSEs for most frames. Furthermore, the proposed method consistently works better than *k-t* FOCUSS for all three sampling patterns studied here.

### E. Reduction Factors

In this experiment, we investigated the reconstruction performance under different reduction factors $R = 2$, 4, 6, 8, and 10. The datasets 1 and 2 were used for evaluation. Fig. 9 shows the ROI images and the errors of the 11th frame in dataset 1 using *k-t* FOCUSS and the proposed method. The visual effects of *k-t* FOCUSS and the proposed method are very close at low reduction factors such as $R = 2$ and 4. However, when the *k*-space data is undersampled further, the proposed method shows superiority in artifacts suppression. Even at a high reduction factor of $R = 10$, the result of proposed method is still free of visible aliasing artifacts and the main structures are well preserved. The variations of MSE depicted in Fig. 10 also demonstrate the superiority of the proposed method at high reduction factors.

### F. Robustness to Noise

Because the above datasets were acquired from real MRI scanners, they already have noise in the measurements. To

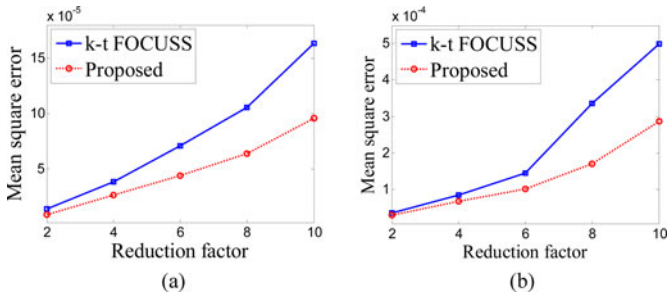


Fig. 10. Comparisons of MSE at different reduction factors. (a) Dataset 1. (b) Dataset 2.

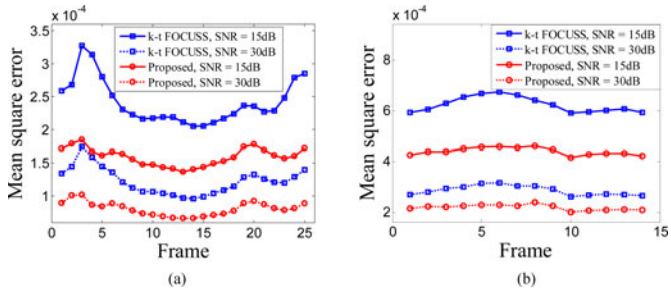


Fig. 12. Comparisons of MSE at SNR = 15 and 30 dB. (a) Dataset 1 ($R = 8$). (b) Dataset 2 ($R = 8$).

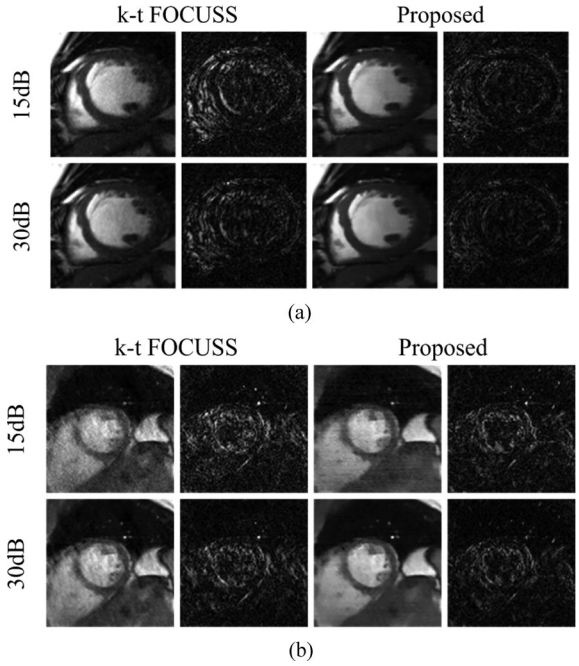


Fig. 11. ROI images and errors at SNR = 15 and 30 dB. (a) Dataset 1 ($R = 8$). (b) Dataset 2 ($R = 8$).

evaluate the robustness of the proposed method to noise, we manually added additional complex Gaussian white noises in the *k*-space data of datasets 1 and 2 with different signal-to-noise ratios (SNR), where the SNR was measured as the ratio of energy between the signal (original data) and the (added) noise. The same set of values was used for the parameters in the proposed method. The results of *k-t* FOCUSS are also presented for comparison. We selected proper parameters for *k-t* FOCUSS to get the lowest MSE.

Fig. 11 shows the ROI images and errors of the 11th frame for dataset 1 and the 12th frame for dataset 2 with a reduction factor $R = 8$. The image qualities of both methods degrade as the SNR decreases. However, the proposed method still obtains acceptable results with less noise and clearer contours than *k-t* FOCUSS even at a low SNR (15 dB). The quality degradation of proposed method is also not as obvious as that of *k-t* FOCUSS. Fig. 12 depicts the frame-by-frame MSEs at SNR = 30 dB and SNR = 15 dB. It is seen that the proposed method outperforms *k-t* FOCUSS at all SNRs. This experiment has shown that the proposed method is robust to noise and has superior denoising ability to *k-t* FOCUSS.

### G. Parameter Evaluation

We evaluated the sensitivity of the proposed method to major parameters including the 3-D DL weighing $(\lambda_1)$, 3-D TV weighting $(\lambda_2)$, patch size $(n_f \times n_p \times n_t)$, the dictionary overcompleteness $(\eta)$, the sparsity constraint $(K)$, and the temporal difference weighting $(\beta_t)$. We used dataset 1 with $R = 8$ as an example for this evaluation. In each experiment, we only changed the values of one parameter and kept other parameters fixed at their default values unless further noted.

We first evaluated the effects of the weighting parameters $\lambda_1$ and $\lambda_2$. We set the penalty parameter $\rho$ to $50\lambda_2$. Fig. 13(a) shows the MSE versus $\lambda_2$ curves under $\lambda_1 = 0.1$, 0.01, and 0.001. The curve of $\lambda_1 = 0.1$ shows much higher MSE than that of smaller $\lambda_1$ values. The MSE trend suggests that the best $\lambda_1$ is $0.01 \leq \lambda_1 \leq 0.001$. Within the range, the U-shaped MSE curve indicates $10^{-5} \leq \lambda_2 \leq 10^{-4}$ to be optimal. The problem is overly or under regularized for $\lambda_1$ and $\lambda_2$ outside the ranges. The rather flat curve also shows the proposed algorithm is not sensitive to the regularization parameters to some extent.

Fig. 13(b) plots the MSEs versus various patch size. Note that the horizontal axis ticks are labeled on the format of "*spatial size* $\times$ *temporal size*." The patch sizes from $3^2 \times 3$ to $8^2 \times 8$ indicate that $n_f \times n_p \times n_t$ varies from $3 \times 3 \times 3$ to $8 \times 8 \times 8$. The patch size $4^2 \times 1$ means that the dictionary is 2-D and the patch size $1^2 \times 25$ implies a 1-D dictionary. For 3-D dictionaries, the MSE is improved with an increase in patch size from 3 to 5, which suggests that small patches cannot express the special features well. When the block size increases from 5 to 8, the reconstruction quality degrades. The reason is probably that larger patches would require dictionaries with more redundancy and our fixed redundancy fails to represent the large patches accurately. It is also seen that the 1-D and 2-D dictionary models lead to higher MSE than 3-D model. The MSE difference between patch sizes 4 and 5 is very small, but the computational time for the latter size is much higher. Therefore, we chose $4 \times 4 \times 4$ as the default patch size.

The redundancy of a dictionary provides more freedom for sparse representation. The increase of the number of atoms generally improves the capability of capturing spatiotemporal details and thus leads to better spatial and temporal resolutions. As

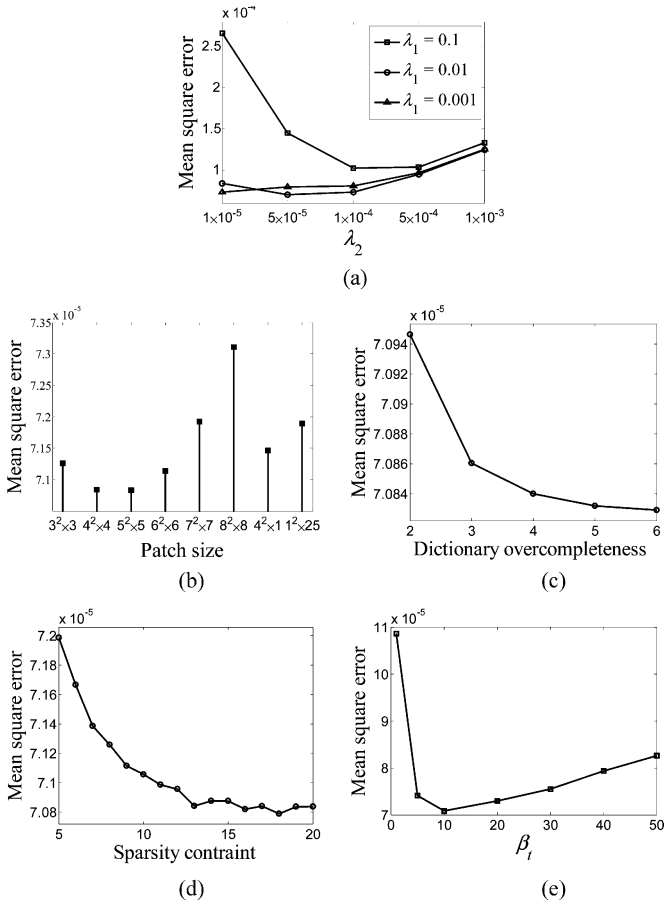


Fig. 13. Performance evaluation with different parameters. (a) MSE versus $\lambda_2$ under different $\lambda_1$. (b) MSE versus patch size. (c) MSE versus the overcompleteness of the dictionary. (d) MSE versus sparsity constraint in D. (e) MSE versus weighting parameter $\beta_t$.

shown in Fig. 13(c), the performance is improved as the redundancy factor increases from 2 to 6. Theoretically, the number of atoms is bounded by the available amount of training patches. However, it is seen that the MSE change from 4 to 6 is rather small, which implies limited improvements for larger overcompletenesses. Considering the computational burden, we chose 4 as the default value.

Fig. 13(d) presents the MSE versus the sparsity constraint. The MSE decreases rapidly as the sparsity threshold is increased from 5 to 13. The reason is that some useful information will be lost if the threshold is too low. When the sparsity level keeps increasing, the MSE becomes more stable and we select 15 as the default value.

The 3-D TV in (15) is able to control the weightings in different directions. We fixed $\beta_x = \beta_y = 1$, changed $\beta_t$ from 1 to 50, and plotted the MSE curve in Fig. 13(e). The performance is initially improved as $\beta_t$ increases. The MSE reaches the minimum at $\beta_t = 10$ which is adopted as the default value. This is because in cardiac cine imaging, only the cardiac region shows notable motions, while temporal variations of other regions are small. The temporal difference images between adjacent frames are sparser (concentrated in the cardiac region) than the spatial difference images, and thus should be weighted more ($\beta_t > \beta_x$, $\beta_y$). However, when $\beta_t$ becomes too large, the temporal profiles are overly smoothed.

These results demonstrate that the proposed algorithm is not overly sensitive to variations of these selected parameters. Although the values of each parameter are set the same in different datasets, we are aware that optimizing parameters for each individual dataset may accelerate the convergence and achieve better performance. However, how to optimize these parameters is still an open problem which is beyond the scope of this paper. It is also worthwhile to note that the newly developed nonparametric Bayesian approach can help mitigate the requirement of prior information in DL [68], but the high computational complexity prevents this approach from being employed in our method at this time.

## V. Discussion

Employing a dictionary to sparsify the 3-D volumes in the image sequence has been presented in [48]. Our approach is different in two aspects. First and foremost, the 3-D volume in [48] comprises a set of image patches from all time frames and the atoms are further enforced to be of low rank. DL with such a large volume size leads to very high computational complexity. In addition, the temporal variations over such a long period of time might be difficult to learn. The proposed 3-D patch representation in this paper should address the issues where a portion of the total frames are extracted to construct the 3-D overlapping patches for DL. Second, although both the proposed model (14) and the formulation in [48] combine 3-D DL with an additional sparsity constraint, our algorithm is derived in a distinct manner. The proposed approach is based on variable splitting and ADMM. Even though the problem is nonconvex, the algorithm still achieves promising results and shows fast convergence rate.

Various experiments have shown the superiority of the proposed method to *k-t* FOCUSS in terms of reconstruction quality. The improvements are enabled by the framework of dictionary-based sparse representation. *k-t* FOCUSS exploits only the temporal correlations but not the spatial structures, resulting in aliasing artifacts along the spatial direction. Furthermore, the predetermined sparsifying transform in *k-t* FOCUSS, namely the temporal Fourier transform, cannot make the cardiac images to be sparse enough so that reconstructions from highly undersampled data show artifacts. In contrast, the proposed model exploits both the spatial and temporal correlations adaptively through the learned dictionary, and thereby the cardiac images can be made sufficiently sparse to be reconstructed from highly reduced data.

Although the proposed method was developed for the single-coil case, we are aware of some parallel dynamic imaging methods, such as paralleled *k-t* FOCUSS [22] and *k-t* SPARSE-SENSE [29]. Parallel imaging can be incorporated into the proposed method by changing the Fourier encoding matrix to sensitivity encoding matrix. Because acquiring accurate coil sensitivities for each time frame can be challenging, we will investigate incorporating parallel imaging in our future study.

## VI. Conclusion

In this paper, we proposed a new CS reconstruction method for dynamic cardiac cine MRI based on a 3-D DL model. The model adopts an adaptively trained spatiotemporal dictionary for sparse representations. The image sequence is divided into 3-D overlapping patches which are enforced to be sparsely expressed over the learned dictionary. We also developed an efficient algorithm based on variable splitting and ADMM. The proposed method is validated using five cardiac cine data. Simulation results demonstrate that the proposed method outperforms *k-t* FOCUSS in terms of artifact removal and image fidelity, especially in high-reduction factor cases. Numerous experiments under different parameter settings show that the proposed method is robust to noise and is not sensitive to parameter changes.

## Acknowledgment

The authors would like to thank Dr. J. C. Ye for making the cardiac cine dataset 1 and the *k-t* FOCUSS code available online. They would also like to thank Dr. E. V. R. Dibella for providing other cardiac cine datasets and thank Dr. M. Elad and Dr. R. Rubinstein for making the K-SVD and OMP code available online.

## References

[1] J. P. Finn, K. Nael, V. Deshpande, O. Ratib, and G. Laub, "Cardiac MR imaging: State of the technology," *Radiology*, vol. 241, no. 2, pp. 338–354, Nov. 2006.

[2] P. Mansfield, "Multi-planar image formation using NMR spin echoes," *J. Phys. C: Solid State Phys.*, vol. 10, no. 3, pp. L55–L58, Feb. 1977.

[3] P. Mansfield and I. L. Pykett, "Biological and medical imaging by NMR," *J. Magn. Reson.*, vol. 29, no. 2, pp. 355–373, Feb. 1978.

[4] C. B. Ahn, J. H. Kim, and Z. H. Cho, "High-speed spiral-scan echo planar NMR imaging-I," *IEEE Trans. Med. Imag.*, vol. MI-5, no. 1, pp. 2–7, Mar. 1986.

[5] C. H. Meyer, B. S. Hu, D. G. Nishimura, and A. Macovski, "Fast spiral coronary artery imaging," *Magn. Reson. Med.*, vol. 28, no. 2, pp. 202–213, Dec. 1992.

[6] D. K. Sodickson and W. J. Manning, "Simultaneous acquisition of spatial harmonics (SMASH): Fast imaging with radiofrequency coil arrays," *Magn. Reson. Med.*, vol. 38, no. 4, pp. 591–603, Oct. 1997.

[7] K. P. Pruessmann, M. Weiger, M. B. Scheidegger, and P. Boesiger, "SENSE: Sensitivity encoding for fast MRI," *Magn. Reson. Med.*, vol. 42, no. 5, pp. 952–962, Oct. 1999.

[8] M. A. Griswold, P. M. Jakob, R. M. Heidemann, M. Nittka, V. Jellus, J. Wang, B. Kiefer, and A. Haase, "Generalized autocalibrating partially parallel acquisitions (GRAPPA)," *Magn. Reson. Med.*, vol. 47, no. 6, pp. 1202–1210, Jun. 2002.

[9] L. Ying and Z.-P. Liang, "Parallel MRI using phased array coils," *IEEE Signal Process. Mag.*, vol. 27, no. 4, pp. 90–98, Jul. 2010.

[10] D. A. Feinberg, J. D. Hale, J. C. Watts, L. Kaufman, and A. Mark, "Halving MR imaging time by conjugation: Demonstration at 3.5 kG," *Radiology*, vol. 161, no. 2, pp. 527–531, Nov. 1986.

[11] D. C. Noll, D. G. Nishimura, and A. Macovski, "Homodyne detection in magnetic resonance imaging," *IEEE Trans. Med. Imag.*, vol. 10, no. 2, pp. 154–163, Jun. 1991.

[12] Z.-P. Liang and P. C. Lauterbur, "An efficient method for dynamic magnetic resonance imaging," *IEEE Trans. Med. Imag.*, vol. 13, no. 4, pp. 677–686, Dec. 1994.

[13] B. Madore, G. H. Glover, and N. J. Pelc, "Unaliasing by Fourier-encoding the overlaps using the temporal dimension (UNFOLD), applied to cardiac imaging and fMRI," *Magn. Reson. Med.*, vol. 42, pp. 813–828, Nov. 1999.

[14] S. J. Riederer, T. Tasciyan, F. Farzaneh, J. N. Lee, R. C. Wright, and R. J. Herfkens, "MR fluoroscopy: Technical feasibility," *Magn. Reson. Med.*, vol. 8, no. 1, pp. 1–15, Sep. 1988.

[15] Z.-P. Liang, H. Jiang, C. P. Hess, and P. C. Lauterbur, "Dynamic imaging by model estimation," *Int. J. Imag. Syst. Technol.*, vol. 8, no. 6, pp. 551–557, Dec. 1997.

[16] Z.-P. Liang, "Spatiotemporal imaging with partially separable functions," in *Proc. IEEE Int. Symp. Biomed. Imag.*, Arlington, TX, USA, 2007, pp. 988–991.

[17] B. Zhao, J. P. Haldar, A. G. Christodoulou, and Z.-P. Liang, "Image reconstruction from highly undersampled (*k, t*)-space data with joint partial separability and sparsity constraints," *IEEE Trans. Med. Imag.*, vol. 31, no. 9, pp. 1809–1820, Sep. 2012.

[18] J. Tsao, P. Boesiger, and K. P. Pruessmann, "k-t BLAST and *k-t* SENSE: Dynamic MRI with high frame rate exploiting spatiotemporal correlations," *Magn. Reson. Med.*, vol. 50, no. 5, pp. 1031–1042, Nov. 2003.

[19] E. J. Candès, J. K. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.

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

[21] M. Lustig, J. M. Santos, D. L. Donoho, and J. M. Pauly, "*k-t* SPARSE: High frame rate dynamic MRI exploiting spatiotemporal sparsity," in *Proc. 14th Annu. Meet. ISMRM*, Seattle, USA, 2006, p. 2420.

[22] H. Jung, J. C. Ye, and E. Y. Kim, "Improved *k-t* BLAST and *k-t* SENSE using FOCUSS," *Phys. Med. Biol.*, vol. 52, no. 11, pp. 3201–3226, Jun. 2007.

[23] U. Gamper, P. Boesiger, and S. Kozerke, "Compressed sensing in dynamic MRI," *Magn. Reson. Med.*, vol. 59, no. 2, pp. 365–373, Feb. 2008.

[24] A. Bilgin, T. Trouard, M. Altbach, and N. Raghunand, "Three dimensional compressed sensing for dynamic MRI," in *Proc. 16th Annu. Meet. ISMRM*, Toronto, Canada, 2008, p. 337.

[25] J. Ji and T. Lang, "Dynamic MRI with compressed sensing imaging using temporal correlations," in *Proc. IEEE Int. Symp. Biomed. Imag.*, Paris, France, 2008, pp. 1613–1616.

[26] G. Adluru, C. McGann, P. Speier, E. G. Kholmovski, A. Shaaban, and E. V. R. DiBella, "Acquisition and reconstruction of undersampled radial data for myocardial perfusion MRI," *J. Magn. Reson. Imag.*, vol. 29, no. 2, pp. 466–473, Feb. 2009.

[27] L. Chen, M. C. Schabel, and E. V. R. DiBella, "Reconstruction of dynamic contrast enhanced magnetic resonance imaging of the breast with temporal constraints," *Magn. Reson. Imag.*, vol. 28, no. 5, pp. 637–645, Jun. 2010.

[28] H. Jung, K. Sung, K. S. Nayak, E. Y. Kim, and J. C. Ye, "*k-t* FOCUSS: A general compressed sensing framework for high resolution dynamic MRI," *Magn. Reson. Med.*, vol. 61, no. 1, pp. 103–116, Jan. 2009.

[29] R. Otazo, D. Kim, L. Axel, and D. K. Sodickson, "Combination of compressed sensing and parallel imaging for highly accelerated first-pass cardiac perfusion MRI," *Magn. Reson. Med.*, vol. 64, no. 3, pp. 767–776, Sep. 2010.

[30] D. Liang, E. V. R. Dibella, R. R. Chen, and L. Ying, "*k-t* ISD: Dynamic cardiac MR imaging using compressed sensing with iterative support detection," *Magn. Reson. Med.*, vol. 68, no. 1, pp. 41–53, Jul. 2012.

[31] M. Lustig, D. Donoho, and J. M. Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging," *Magn. Reson. Med.*, vol. 58, no. 6, pp. 1182–1195, Dec. 2007.

[32] A. S. Gupta and Z.-P. Liang, "Dynamic imaging by temporal modeling with principle component analysis," in *Proc. 9th Annu. Meet. ISMRM*, Glasgow, Scotland, 2001, p. 10.

[33] R. Rubinstein, A. M. Bruckstein, and M. Elad, "Dictionaries for sparse representation modeling," *Proc. IEEE*, vol. 98, no. 6, pp. 1045–1057, Jun. 2010.

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

[35] R. Rubinstein, T. Peleg, and M. Elad, "Analysis K-SVD: A dictionary learning algorithm for the analysis sparse model," *IEEE Trans. Signal Process.*, vol. 61, no. 3, pp. 661–677, Feb. 2013.

[36] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Trans. Image Process.*, vol. 15, no. 12, pp. 3736–3745, Dec. 2006.

[37] M. Protter and M. Elad, "Image sequence denoising via sparse and redundant representations," *IEEE Trans. Image Process.*, vol. 18, no. 1, pp. 27–36, Jan. 2009.

[38] I. Ramirez, P. Sprechmann, and G. Sapiro, "Classification and clustering via dictionary learning with structured incoherence and shared features," in *Proc. IEEE Comput. Vision Pattern Recog.*, San Francisco, CA, USA, 2010, pp. 3501–3508.

[39] S. Kong and D. Wang, "A dictionary learning approach for classification: Separating the particularity and the commonality," in *Proc. 12th Eur. Conf. Comput. Vision*, Firenze, Italy, 2012, pp. 186–199.

[40] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 210–227, Feb. 2009.

[41] L. Ma, C. Wang, B. Xiao, and W. Zhou, "Sparse representation for face recognition based on discriminative low-rank dictionary learning," in *Proc. IEEE Conf. Comput. Vision Pattern Recog.*, Providence, USA, 2012, pp. 2586–2593.

[42] X. Ye, Y. Chen, and F. Huang, "A novel method and fast algorithm for MR image Reconstruction with significantly undersampled data," *Inverse Problems Imag.*, vol. 4, no. 2, pp. 223–240, May 2010.

[43] S. Ravishankar and Y. Bresler, "MR image reconstruction from highly undersampled k-space data by dictionary learning," *IEEE Trans. Med. Imag.*, vol. 30, no. 5, pp. 1028–1040, May 2011.

[44] Q. Liu, S. Wang, K. Yang, J. Luo, Y. Zhu, and D. Liang, "Highly undersampled magnetic resonance image reconstruction using two-level Bregman method with dictionary updating," *IEEE Trans. Med. Imag.*, vol. 32, no. 7, pp. 1290–1301, Jul. 2013.

[45] M. Doneva, P. Börnert, H. Eggers, C. Stehning, J. Sénégas, and A. Mertins, "Compressed sensing reconstruction for magnetic resonance parameter mapping," *Magn. Reson. Med.*, vol. 64, no. 4, pp. 1114–1120, Oct. 2010.

[46] Y. Wang, Y. Zhou, and L. Ying, "Undersampled dynamic magnetic resonance imaging using patch-based spatiotemporal dictionaries," in *Proc. IEEE Int. Symp. Biomed. Imag.*, San Francisco, CA, USA, 2013, pp. 294–297.

[47] Y. Wang and L. Ying, "Accelerating dynamic MRI using patch-based spatiotemporal dictionaries," in *Proc. 21th Annu. Meet. ISMRM*, Salt Lake City, UT, USA, 2013, p. 3804.

[48] S. P. Awate and E. V. R. DiBella, "Spatiotemporal dictionary learning for undersampled dynamic MRI reconstruction via joint frame-based and dictionary-based sparsity," in *Proc. IEEE Int. Symp. Biomed. Imag.*, Barcelona, Spain, 2012, pp. 318–321.

[49] Y. Song and J. Zhao, "Dynamic MR imaging reconstruction by three-dimensional dictionary learning," in *Proc. 5th Int. Conf. Biomed. Eng. Informat.*, Chongqing, China, 2012, pp. 35–39.

[50] J. Caballero, D. Rueckert, and J. V. Hajnal, "Dictionary learning and time sparsity in dynamic MRI," in *Proc. 15th Med. Image Comput. Comput. Assist. Interv.*, Nice, France, 2012, pp. 256–263.

[51] J. Caballero, A. N. Price, D. Rueckert, and J. V. Hajnal, "Dictionary based reconstruction of dynamic complex MRI data," in *Proc. 21th Annu. Meet. ISMRM*, Salt Lake City, UT, USA, 2013, p. 3806.

[52] E. J. Candès and T. Tao, "Decoding by linear programming," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.

[53] E. J. Candès, "The restricted isometry property and its implications for compressed sensing," *Comptes Rendus Mathematique*, vol. 346, no. 9, pp. 589–592, May 2008.

[54] D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via $\ell^1$ minimization," *Proc. Nat. Acad. Sci. USA*, vol. 100, no. 5, pp. 2197–2202, Mar. 2003.

[55] E. J. Candès and J. Romberg, "Sparsity and incoherence in compressive sampling," *Inverse Problem*, vol. 23, no. 3, pp. 969–985, Apr. 2007.

[56] J. A. Tropp and S. J. Wright, "Computational methods for sparse solution of linear inverse problems," *Proc. IEEE*, vol. 98, no. 6, pp. 948–958, Jun. 2010.

[57] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *Proc. 27th Annu. Asilomar Conf. Signals Syst. Comput.*, Pacific Grove, CA, USA, 1993, vol. 1, pp. 40–44.

[58] T. Blumensath and M. Davies, "Iterative hard thresholding for compressive sensing," *Appl. Comput. Harmon. Anal.*, vol. 27, no. 3, pp. 265–274, Nov. 2009.

[59] R. Chartrand, "Exact reconstruction of sparse signals via nonconvex minimization," *IEEE Signal Process. Lett.*, vol. 14, no. 10, pp. 707–710, Oct. 2007.

[60] D. Liang and L. Ying, "A hybrid L0-L1 minimization algorithm for compressed sensing MRI," in *Proc. 18th Annu. Meet. ISMRM*, Stockholm, Sweden, 2010, p. 4869.

[61] K. Engan, S. O. Aase, and J. Hakon Husoy, "Method of optimal directions for frame design," in *Proc. Int. Conf. Acoust., Speech, Signal Process.*, Phoenix, AZ, USA, 1999, vol. 5, pp. 2443–2446.

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

[63] S. H. Chan, R. Khoshabeh, K. B. Gibson, P. E. Gill, and T. Q. Nguyen, "An augmented Lagrangian method for total variation video restoration," *IEEE Trans. Image Process.*, vol. 20, no. 11, pp. 3097–3111, Nov. 2011.

[64] T. Goldstein and S. Osher, "The split Bregman method for l1-regularized problems," *SIAM J. Imag. Sci.*, vol. 2, no. 2, pp. 323–343, Apr. 2009.

[65] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends in Mach. Learning*, vol. 3, no. 1, pp. 1–122, Jul. 2011.

[66] M. V. Afonso, J. M. Bioucas-Dias, and M. A. T. Figueiredo, "An augmented Lagrangian approach to the constrained optimization formulation of imaging inverse problems," *IEEE Trans. Image Process.*, vol. 20, no. 3, pp. 681–695, Mar. 2011.

[67] A. Matakos, S. Ramani, and J. A. Fessler, "Accelerated edge-preserving image restoration without boundary artifacts," *IEEE Trans. Image Process.*, vol. 22, no. 5, pp. 2019–2029, May 2013.

[68] M. Zhou, H. Chen, J. Paisley, L. Ren, L. Li, Z. Xing, D. Dunson, G. Sapiro, and L. Carin, "Nonparametric Bayesian dictionary learning for analysis of noisy and incomplete images," *IEEE Trans. Image Process.*, vol. 21, no. 1, pp. 130–144, Jan. 2012.

Authors' photographs and biographies not available at the time of publication.