Frontiers of Information Technology & Electronic Engineering
www.zju.edu.cn/jzus; engineering.cae.cn; www.springerlink.com
ISSN 2095-9184 (print); ISSN 2095-9230 (online)
E-mail: jzus@zju.edu.cn



Multi-scale UDCT dictionary learning based highly undersampled MR image reconstruction using patch-based constraint splitting augmented Lagrangian shrinkage algorithm*

Min YUAN, Bing-xin YANG, Yi-de MA†‡, Jiu-wen ZHANG, Fu-xiang LU, Tong-feng ZHANG

(*School of Information Science & Engineering, Lanzhou University, Lanzhou 730000, China*)

†E-mail: ydma01@126.com

Received Dec. 8, 2014; Revision accepted Oct. 9, 2015; Crosschecked Oct. 12, 2015

Abstract: Recently, dictionary learning (DL) based methods have been introduced to compressed sensing magnetic resonance imaging (CS-MRI), which outperforms pre-defined analytic sparse priors. However, single-scale trained dictionary directly from image patches is incapable of representing image features from multi-scale, multi-directional perspective, which influences the reconstruction performance. In this paper, incorporating the superior multi-scale properties of uniform discrete curvelet transform (UDCT) with the data matching adaptability of trained dictionaries, we propose a flexible sparsity framework to allow sparser representation and prominent hierarchical essential features capture for magnetic resonance (MR) images. Multi-scale decomposition is implemented by using UDCT due to its prominent properties of lower redundancy ratio, hierarchical data structure, and ease of implementation. Each sub-dictionary of different sub-bands is trained independently to form the multi-scale dictionaries. Corresponding to this brand-new sparsity model, we modify the constraint splitting augmented Lagrangian shrinkage algorithm (C-SALSA) as patch-based C-SALSA (PB C-SALSA) to solve the constraint optimization problem of regularized image reconstruction. Experimental results demonstrate that the trained sub-dictionaries at different scales, enforcing sparsity at multiple scales, can then be efficiently used for MRI reconstruction to obtain satisfactory results with further reduced undersampling rate. Multi-scale UDCT dictionaries potentially outperform both single-scale trained dictionaries and multi-scale analytic transforms. Our proposed sparsity model achieves sparser representation for reconstructed data, which results in fast convergence of reconstruction exploiting PB C-SALSA. Simulation results demonstrate that the proposed method outperforms conventional CS-MRI methods in maintaining intrinsic properties, eliminating aliasing, reducing unexpected artifacts, and removing noise. It can achieve comparable performance of reconstruction with the state-of-the-art methods even under substantially high undersampling factors.

Key words: Compressed sensing (CS), Magnetic resonance imaging (MRI), Uniform discrete curvelet transform (UDCT), Multi-scale dictionary learning (MSDL), Patch-based constraint splitting augmented Lagrangian shrinkage algorithm (PB C-SALSA)

doi:10.1631/FITEE.1400423 **Document code:** A **CLC number:** TN911

1 Introduction

Magnetic resonance imaging (MRI) is a widely used noninvasive imaging modality for clinical diagnosis. However, low speed in traditional magnetic resonance (MR) data acquisition due to physical and physiological constraints remains a great challenge for clinical applications. One method to speed up data acquisition is the *k*-space undersampling. However, undersampling violates the Nyquist sampling criterion,

‡ Corresponding author

* Project supported by the National Natural Science Foundation of China (Nos. 61175012 and 61201422), the Natural Science Foundation of Gansu Province of China (No. 1208RJ-ZA265), the Specialized Research Fund for the Doctoral Program of Higher Education of China (No. 2011021111-0026), and the Fundamental Research Funds for the Central Universities of China (Nos. lzujbky-2015-108 and lzujbky-2015-197)

ORCID: Min YUAN, http://orcid.org/0000-0001-7855-8678

© Zhejiang University and Springer-Verlag Berlin Heidelberg 2015

leading to aliasing artifacts. Besides, noise due to coils and eddy currents of human body influence the accuracy of *k*-space data. The above two factors blur MR image edges containing significant information in pathological diagnosis. Hence, undersampling MR image reconstruction without degrading image quality has become a focus in the field of medical imaging in recent years.

Compressed sensing (CS) (Candes and Donoho, 2004; Candes *et al*., 2006a; 2006b; Baraniuk, 2007) is a novel signal acquisition and compression technique that enables accurate signal recovery from dramatically few measurements only if the signal is sparse or can be sparsely represented in a sparsifying transform domain. CS has shown great potential in reducing the data acquisition time for MRI. Compressed sensing MRI (CS-MRI) (Lustig *et al*., 2007; 2008) allows high-quality reconstruction from highly undersampled *k*-space data by solving a constrained minimization problem with nonlinear optimization methods on the premise of the sparsity of images. CS-MRI has been a topic of great interest to the MR reconstruction community. A leading approach using Daubechies wavelet transform and total variation (TV) (Rudin *et al*., 1992) as the sparsity method and the nonlinear conjugate gradient descent algorithm for reconstruction (denoted as LDP) (Lustig *et al*., 2007; 2008) has verified the efficiency of CS in rapid MRI.

Sparsifying transform plays a key role in CS-MRI. The performance of reconstruction from highly undersampled *k*-space data relies greatly on the chosen sparsity prior. Numerous predefined analytical sparsity approaches have been exploited in CS-MRI, such as wavelet transform (Lustig *et al*., 2007), TV (Lustig *et al*., 2007; 2008; Huang *et al*., 2011), contourlet (Gho *et al*., 2010), and multi-scale geometric analysis (MGA) (Kim *et al*., 2009; Qu *et al*., 2010; Zhu *et al*., 2013). The structured sparsity, such as the Gaussian scale mixture (GSM) model (Kim *et al*., 2012) and wavelet tree sparsity (Chen and Huang, 2014), has been introduced to CS-MRI reconstruction. However, the predefined sparsifying transforms lack the adaptability to various images, and thus could not provide a sufficient sparse representation for the reconstructed images. In general, the performance of CS-MRI reconstruction with nonadaptive sparsifying transforms at low undersampling rates is not good enough. Consequently, the loss of important edge details and artifacts can be shown in reconstructed images from undersampled *k*-space data due to the lack of local structures. These artifacts may be viewed as meaningful edges and possibly mislead the diagnosis. In contrast, learning adaptive transforms in CS-MRI can lead to lower reconstruction errors and superior image quality because the dictionary learned from the training sample can effectively capture local image features and provide a sparser representation, and can be finely fitted to data (Lewicki and Sejnowski, 2000; Rauhut *et al*., 2008; Elad, 2010).

Dictionary learning (DL) (Elad, 2010) from intermediate reconstruction or fully sampled reference images has also been introduced for CS-MRI (Ravishankar and Bresler, 2011; Qu *et al*., 2012; Ning *et al*., 2013). The trained dictionary is expected to provide sparser representation of MR images compared with the general sparsifying transforms, thus improving the reconstruction.

Ravishankar and Bresler (2011) proposed a novel framework, called dictionary learning MRI (DLMRI), to learn the sparsifying transform (dictionary) adaptively and reconstruct the image simultaneously from highly undersampled *k*-space data. However, in DLMRI, sparse dictionaries are trained in the image domain at a single scale. MR images usually contain some anisotropic features at different scales. These intrinsic features need to be retained for image reconstruction. A dictionary trained from image patches can fit the data finely, but it fails to represent these characteristics from multiple scales in a hierarchical structure.

Qu *et al*. (2012) proposed the method of patch-based directional wavelets (PBDW), which trains geometric directions from undersampled data. However, PBDW cannot sparsify the smooth regions better than the conventional sparsifying transform, which leads to artifacts that influence geometric directions. Afterwards, in PBDWS (Ning *et al*., 2013), the performance of PBDW-based undersampled MR image reconstruction has been improved by minimizing the l_0-norm of transform coefficients and extending PBDW into a two-dimensional (2D) shift-invariant discrete wavelet domain to enhance the ability of transform on sparsifying piecewise smooth image features. Qu *et al*. (2014) designed a patch-based nonlocal operator (PANO) to sparsify MR images by using the similarity of image patches. The

operator also leads to optimized sparse representation of images to be reconstructed, whereas the availability of the guided image and how the gridding process affects PANO imaging with non-Cartesian sampling remain to be carefully analyzed.

Hence, exploiting a more sparsifying method for image representation to break the bottleneck of undersampling rate is crucial. The experience with the preconstructed transforms tells us that it is generally beneficial to process the signals in several different scales, and then to operate on each scale separately. This leads us to another approach for seeking proper dictionaries, which overcomes these limitations by adopting both multi-scale and learning points of view. By integrating the advantages of the above two methods and merging a learning procedure on top of an existing multi-scale transform, a novel framework of multi-scale dictionary learning (MSDL) (Rubinstein *et al*. 2010; Ophir *et al*., 2011) has been built. MSDL combines the multi-scale property of MGA and the superior data adaptability of trained dictionaries. Inspired by this, using the superiority of uniform discrete curvelet transform (UDCT) (Nguyen and Chauris, 2010) and patch-based methodologies, in this paper, we explore the incorporation of adaptive dictionary learning with UDCT representation to introduce a more flexible framework for learning multi-scale sparse representations of MR images with overcomplete dictionaries.

Our proposed CS-MRI algorithm is termed 'UDCT DL patch-based C-SALSA', UDPC for short. UDPC contains two stages. A simple and easily implemented zero-filling reconstruction image is chosen as the intermediate reconstruction for dictionary training. Considering that the realistic MR data generally contain phase information, the real and imaginary parts of the data are treated independently in UDPC. In the first stage, the real and imaginary parts of the initialized reconstruction image are decomposed into multi-scale and multi-directional subbands applying UDCT, respectively. Each subdictionary of multi-scale dictionaries is trained independently using the K-singular value decomposition (K-SVD) algorithm (Aharon *et al*., 2006; Elad and Aharon, 2006) from maximum overlapped patches of the corresponding UDCT sub-band coefficients. In the second stage, corresponding to blocking of MSDL in the UDCT domain, we extend the constraint splitting augmented Lagrangian shrinkage algorithm (C-SALSA) (Afonso *et al*., 2011) to fit patch-based sub-dictionaries, termed 'patch-based C-SALSA' (PB C-SALSA), to update the solution repeatedly until reaching convergence. The previously obtained reconstruction image will serve as the intermediate reconstruction image for the next iteration in multi-scale dictionary training. These two stages are performed alternately until reaching convergence.

The multi-scale UDCT dictionaries trained by the proposed sparsifying model can give rise to the more prominent sparsity for the particular image instance at multiple scales compared with the state-of-the-art methods, thereby leading to accurate capturing of the intrinsic characteristics of MR images and substantially improved reconstruction performance for CS-MRI. These dictionaries can then be used efficiently for MRI reconstruction to obtain satisfactory results with further reduced undersampling rate, potentially outperforming both single-scale trained dictionaries and multi-scale analytic transforms. In addition, the extended PB C-SALSA with high efficiency and stability leads to fast convergence for the implementation of MR image reconstruction.

2 Background and related work

2.1 CS-MRI

Define $\boldsymbol{x}\in\mathbb{C}^P$ as the vector of the 2D image signal to be reconstructed and k-space measurements $\boldsymbol{y}=\boldsymbol{F}_{\mathrm{u}}\boldsymbol{x}$, where $\boldsymbol{F}_{\mathrm{u}}\in\mathbb{C}^{m\times P}$ represents the undersampled Fourier encoding matrix. In CS-MRI, assume $\boldsymbol{\Psi}\in\mathbb{C}^{T\times P}$ represents the analytical sparsifying transform or a set of signals learned from image patches. Sparse representation is defined as $\boldsymbol{\alpha}=\boldsymbol{\Psi x}$, where $\boldsymbol{\alpha}$ is assumed to be sparse corresponding to image $\boldsymbol{x}$ in domain $\boldsymbol{\Psi}$. Reconstructing the unknown MR image $\boldsymbol{x}$ from measurements $\boldsymbol{y}$ using CS is to solve the CS-MRI optimization problem, which is denoted as

$$\min_{x} \|\boldsymbol{\Psi x}\|_0 \text{ s.t. } \boldsymbol{F}_{\mathrm{u}}\boldsymbol{x}=\boldsymbol{y}. \tag{1}$$

However, this l_0 problem is generally nonconvex and NP-hard (Candes and Donoho, 2004; Candes

et al., 2006a; 2006b; Donoho, 2006). Under certain conditions, this l_0-norm problem can be approximately replaced by the convex relaxation, l_1-norm optimization, which simplifies the solution process of Eq. (1) and promotes sparsity. Besides, in the process of *k*-space data acquisition, random noise is unavoidable. The minimization problem in Eq. (1) turns into

$$\min_{\boldsymbol{x}} \| \boldsymbol{\Psi x} \|_1 \text{ s.t. } \| \boldsymbol{F}_{\mathrm{u}} \boldsymbol{x} - \boldsymbol{y} \|_2 \leq \varepsilon, \tag{2}$$

where ε is a parameter which depends upon the added noise variance. By merging the constraint term into objective function (2), the formula turns out to be

$$\min_{\boldsymbol{x}} \left(\frac{1}{2} \| \boldsymbol{F}_{\mathrm{u}} \boldsymbol{x} - \boldsymbol{y} \|_2^2 + \lambda \| \boldsymbol{\Psi x} \|_1 \right), \tag{3}$$

where the Lagrangian multiplier λ>0 controls the tradeoff between solution sparsity and data fidelity. In Eq. (3), the error term is used to constrain the consistency of the reconstructed image with *k*-space data, and the sparse constraint term is used to guarantee the sparsity in the transform domain.

As pointed out by Afonso *et al*. (2011), parameter ε in Eq. (2) has a straightforward meaning, which is proportional to the noise standard deviation, and is much easier to set than parameter λ in Eq. (3). Consequently, in this work, we focus directly on the constrained problems (1) and (2) by a fast algorithm.

2.2 Uniform discrete curvelet transform

UDCT (Nguyen and Chauris, 2010) is a novel mathematical and computational tool for multi-resolution data representation and an innovative implementation of the discrete curvelet transform, which uses the ideas of fast Fourier transform (FFT) based discrete curvelet transform and filter-bank based contourlet transform. The discrete curvelet functions are defined by a parameterized family of smooth windowed functions that satisfy two conditions: they are 2π periodic and their squares form a partition of unity, and the centers of the curvelet functions at each resolution are positioned on a uniform lattice. UDCT is implemented by the FFT algorithm but designed as a multi-resolution filter-bank with the advantages of the two methods.

Compared with other directional, discrete, and nonadaptive transforms, UDCT provides a flexible instead of fixed number of directions at each level to accurately capture various directional geometrical structures of the image. UDCT has several advantages over existing transforms in practical applications, such as lower redundancy ratio, hierarchical data structure, and ease of implementation. These make UDCT very practical in many applications. Furthermore, its shift-invariance in the energy sense is significant in image analysis and representation. Fig. 1 illustrates some effective atoms learned from different UDCT coefficient sub-bands of some scales after UDCT operating on the image. For one sub-dictionary, all coefficients are set to zeros except one in terms of this sub-dictionary, to visualize a single effective atom of it. Passing the result of multiplying such a coefficient set by the learned multi-scale dictionary, through the uniform discrete curvelet synthesis operation, exhibits a visualization of a single 'effective' atom in the image domain, demonstrating that the atoms are localized and from different scales, possess clear directionality, and are adapted to the training data.

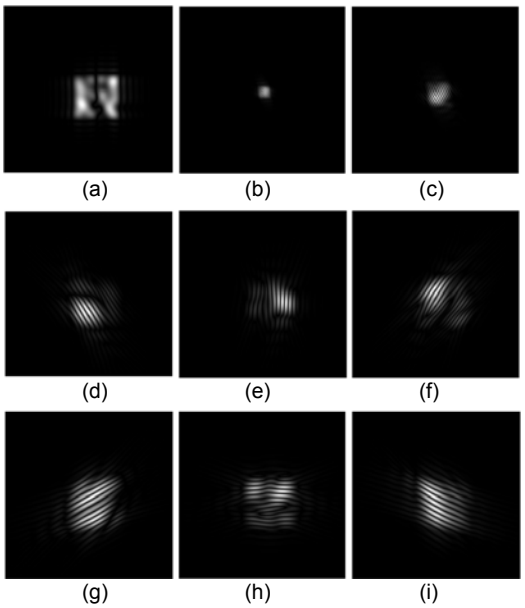


Fig. 1 Visualization of some effective atoms from different levels/bands trained on an undersampling T2-weighted image of the brain using a four-level UDCT

A separate sub-dictionary was trained for each band. Atoms came from the approximation band (a), the first direction of the second level (b), the first direction of the third level (c), and six directions of the fourth level (d)–(i)

In contrast, UDCT is more effective in approximating piecewise smooth images containing rich geometric information such as edges, corners, and spikes, which is of vital importance for CS-MRI applications. UDCT has very similar asymptotic approximation properties to the curvelets (Candes and Donoho, 2004). To make the statement of optimal sparsity and approximation property more rigorous, it is useful to quantify the approximation performance from the point of view of approximation theory. The asymptotic convergence is actually the correct optimal behavior for approximating general smooth objects having discontinuities along piecewise C^2 curves. Denoting $\hat{\boldsymbol{f}}_M$ as the approximation of an image $\boldsymbol{f}$ by the M largest transform coefficients in the corresponding representation, the resulting approximation error (in l_2-norm square) is $\| \boldsymbol{f} - \hat{\boldsymbol{f}}_M \|_2^2$. It is very helpful to achieve the best asymptotic decay rate for this error in applications. Let C^2 be the space of functions that are twice continuously differentiable. If image $\boldsymbol{f}$ is C^2 everywhere away from edge curves that are piecewise C^2, the best M-term asymptotic approximation error using curvelets has a decay rate of $O((\log M)^3 M^{-2})$ (Candes and Donoho, 2004), which is essentially optimal in representing 2D image $\boldsymbol{f}$ which is piecewise C^2 except for discontinuities along C^2 curves, and greatly outperforms that of wavelet approximations only with the decay rate of $O(M^{-1})$ (Mallat, 2008). The error decay rate using UDCT is close to the theoretical optimal approximation, where the error decays as M^{-2} (Donoho, 2001). In this sense, the UDCT representation provides an optimally sparse representation of objects with singularities along piecewise C^2 edges. The transform has the same order of complexity as FFT. Consequently, we consider the applications of UDCT using these advantages for multi-resolution data representation. UDCT is well suited to the problem of CS-MRI.

2.3 Dictionary learning

Dictionary learning can capture fine structures and details effectively for the image reconstruction problem. In addition, superior adaptability of trained dictionaries to various anisotropic image features could be beneficial to represent the images sparsely. Dictionaries learned from trained samples provide an adaptive sparse representation for images. For a complex-valued image $\boldsymbol{x} \in \mathbb{C}^P$, let $\boldsymbol{R}_{ij} \in \mathbb{C}^{n \times P}$ be the operator that extracts patch $\boldsymbol{x}_{ij} \in \mathbb{C}^n$ (image patch size $\sqrt{n} \times \sqrt{n}$, and indexed by (i, j), the location of its top-left corner in the image) as a column vector from $\boldsymbol{x}$ denoted as $\boldsymbol{x}_{ij}=\boldsymbol{R}_{ij}\boldsymbol{x}$. $\sum_{ij} \boldsymbol{R}_{ij}^{\mathrm{T}} \boldsymbol{R}_{ij} \in \mathbb{C}^{P \times P}$ represents a diagonal matrix in which the diagonal elements correspond to the pixel locations of $\boldsymbol{x}$. Each image patch $(\boldsymbol{x}_{ij})$ can be sparsely represented by a learned dictionary $\boldsymbol{D} \in \mathbb{C}^{n \times K}$, i.e., $\boldsymbol{x}_{ij}=\boldsymbol{D}\boldsymbol{\alpha}_{ij}$ ($\boldsymbol{\alpha}_{ij}$ is a sparse coefficient vector of $\boldsymbol{x}_{ij}$ with respect to dictionary $\boldsymbol{D}$). For all the patches, the dictionary learning model can be written as

$$\begin{aligned} &\min_{\boldsymbol{D},\Gamma} \sum_{i,j} \| \boldsymbol{R}_{ij}\boldsymbol{x} - \boldsymbol{D}\boldsymbol{\alpha}_{ij} \|_2^2 \\ &\text{s.t.} \quad \| \boldsymbol{\alpha}_{ij} \|_0 \le T_0, \quad \forall i, j, \end{aligned} \tag{4}$$

where Γ is the sparse representation set $\{\boldsymbol{\alpha}_{ij}\}_{ij}$ of all training patches, and T_0 the required sparsity level. The l_0 quasi norm is used to constrain the sparsity of each patch representation. Sparse coding is the process to find a sparse coefficient vector $\boldsymbol{\alpha}$ with a given initial dictionary $\boldsymbol{D}$. To tackle this DL problem, numerous algorithms have been proposed, concentrating mainly on the alternating process between finding the optimal dictionary and corresponding sparse coding. Particularly, the K-SVD algorithm (Aharon *et al*., 2006), derived from K-means, outperforms others in terms of simplicity, flexibility, and efficiency. In K-SVD, the dictionary $\boldsymbol{D}$ is trained by solving the optimization problem, where data are fitted according to sparsity constraints.

Thus, a possible formula for CS-MRI based on DL from image patches can be denoted as

$$\begin{aligned} &\min_{\boldsymbol{x},\boldsymbol{D},\Gamma} \sum_{i,j} \| \boldsymbol{R}_{ij}\boldsymbol{x} - \boldsymbol{D}\boldsymbol{\alpha}_{ij} \|_2^2 + \delta \| \boldsymbol{F}_{\mathrm{u}}\boldsymbol{x} - \boldsymbol{y} \|_2^2 \\ &\text{s.t.} \quad \| \boldsymbol{\alpha}_{ij} \|_0 \le T_0, \ \forall i, j. \end{aligned} \tag{5}$$

The first term in Eq. (5) is used to capture the sparse approximation of $\boldsymbol{x}$ with respect to dictionary $\boldsymbol{D}$, and the second term is applied to ensure k-space data fidelity. The positive constant δ represents the weight of the fidelity term in the cost function, which is relevant to the noise level, making the formulation more robust to noise.

DLMRI is an appealing framework to tackle Eq. (5), which involves settling an alternating minimization problem of the DL procedure using K-SVD and the reconstruction update process, and achieves a better reconstruction result than other earlier methods. However, trained sparse dictionaries in DLMRI operate on images by considering only their native scales, which fails to represent intrinsic features from multiple scales and multiple directions. Thus, it influences further improvement of reconstruction performance. Furthermore, it is empirical and impractical to determine the optimal reconstruction parameter values in the DL process of DLMRI so as to reach improved reconstruction results for every specified image in each iteration process attributed to the nonconvergence of its applied reconstruction approach. Meanwhile, the large number of iterations between the DL and image update process increases the computational complexity. On the other hand, most experimental validations (except two cases) in Ravishankar *et al*. (2011) were performed using the images downloaded from American Radiology Services, which have been interpolated and compressed. This means that their characteristics are very different from real MR images. Real-valued MR images have discarded phase information in realistic data. The raw MR data generally are complex-valued, containing phase information. The reconstruction results of DLMRI for complex-valued data will be discussed in Section 4.

3 Proposed UDPC algorithm

The predefined, global sparsifying transforms are typically limited in their ability to sparsify signals, which leads to the insufficient sparse representation for images of a certain type without adaptability. Learned dictionary from the image domain at a single scale can adaptively represent information for the particular image instance and finely fit the data, yet it is unable to capture the essence of image in various scales and directions. Multi-scale dictionary learning effectively combines the advantages of these two representations. Consequently, we propose a CS-MRI reconstruction method based on UDCT multi-scale dictionary learning. The framework of sparsifying the images combines the advantages of multi-scale UDCT and dictionary-learning methodology. Considering the computational load and PB C-SALSA, an efficient iterative algorithm with fast convergence is introduced for the implementation of reconstruction.

The flowchart of the proposed UDPC-based MRI reconstruction from undersampled data is shown in Fig. 2. In practice, the obtained *k*-space measurements are always complex-valued data with a phase component. So, we consider separating the real and imaginary parts of *k*-space data, and then handling them seperately. Multi-scale directional analysis is implemented by adopting a multi-level UDCT decomposition. To further sparsify data, we exploit dictionary learning to enhance the sparsity of UDCT coefficients. Overcomplete dictionaries are trained on each sub-band in the UDCT domain. Each sub-dictionary is trained independently adapting to each coefficient sub-band of each level to deliver a significantly sparser image representation superior to multi-resolution analysis alone and single-scale dictionary alone.

We propose the following model to reconstruct $\hat{\boldsymbol{x}}$ in UDPC:

$$\min_{\boldsymbol{x},\boldsymbol{D},\Gamma,\boldsymbol{\Gamma}_t} \sum_{b=1}^{B}\sum_{i,j} \| \boldsymbol{R}_{ij}(\boldsymbol{\Psi x})_b - \boldsymbol{D}_b\boldsymbol{\alpha}_{ij} \|_2^2 + \lambda\sum_{b=1}^{B}\phi((\boldsymbol{D}_b)^{\#}(\boldsymbol{\Psi x})_b)$$
$$\text{s.t. } \| \boldsymbol{\alpha}_{ij} \|_0 \le T_b,\ \| \boldsymbol{F}_{\mathrm{u}}\boldsymbol{x}-\boldsymbol{y} \|_2 \le \varepsilon,\ \forall i,j, \tag{6}$$

where $\boldsymbol{\Psi}$ denotes the UDCT operator applied to image $\boldsymbol{x}$; thus, $(\boldsymbol{\Psi x})_b$ represent different coefficient sub-bands of the decomposition coefficients $\boldsymbol{\Psi x}$ by the UDCT. $\boldsymbol{D}_b$ (b=1, 2, …, B) represent the relevant sub-dictionaries learned from the UDCT coefficient sub-bands, where B is the total number of sub-bands which is related to the number of scales (J) and the number of directions for each scale ($2v_J$), i.e., $B=J\times 2v_J+1$. The set $\{\boldsymbol{D}_b|b=1, 2, \ldots, B\}$ is the learned multi-scale dictionary in the UDCT domain. $(\boldsymbol{D}_b)^{\#}$ denotes the pseudo-inverse of the trained UDCT coefficient sub-band dictionary $\boldsymbol{D}_b$ such that $(\boldsymbol{D}_b)^{\#}=((\boldsymbol{D}_b)^{\mathrm{H}}\boldsymbol{D}_b)^{-1}(\boldsymbol{D}_b)^{\mathrm{H}}$, where $(\boldsymbol{D}_b)^{\mathrm{H}}$ indicates the Hermitian transpose of $\boldsymbol{D}_b$. T_b is the corresponding sparsity level with respect to each sub-dictionary and ϕ is a convex, nonsmooth regularizer, such as TV or l_1-norm. Practical implementation of reconstructing $\hat{\boldsymbol{x}}$ by the proposed model contains two stages, i.e., multi-scale UDCT dictionary learning and MR image

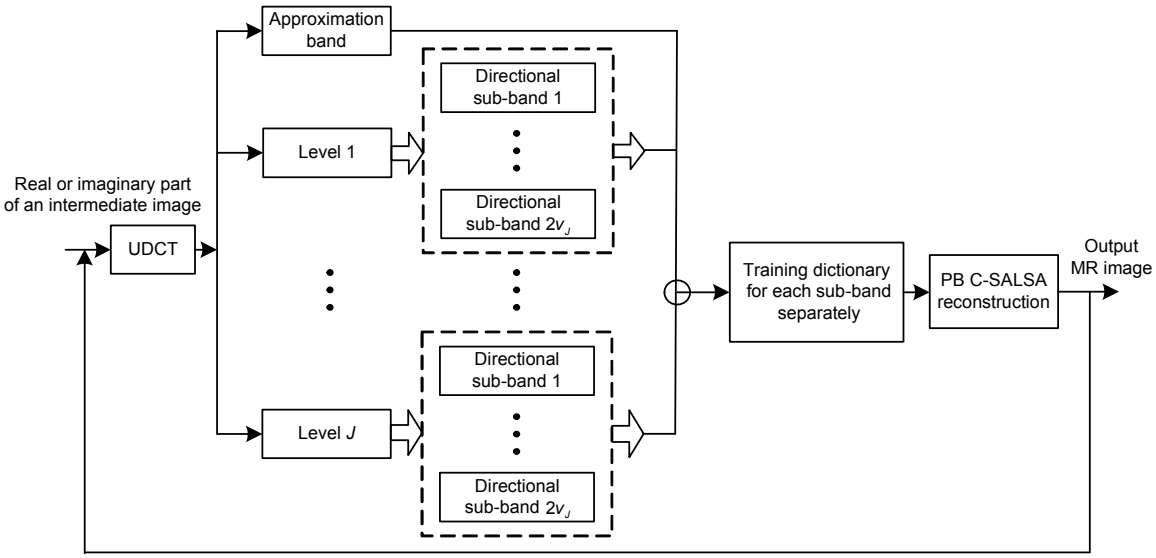


Fig. 2 Flowchart of the proposed UDPC-based MRI reconstruction

reconstruction (solving the constrained optimization problem by regularization).

3.1 Multi-scale dictionary learning based on UDCT

In this stage, the zero-filling reconstruction image $\boldsymbol{x}_0 = \boldsymbol{F}_{\mathrm{u}}^{\mathrm{H}}\boldsymbol{y}$ is adopted as the intermediate image to initialize the reconstructed MR image. The real and imaginary parts of the initial image are decomposed by the UDCT into one lowpass, three directional scales (J=3), and $2v_j$=6 (j=1, 2, 3) directions in each scale, respectively. Thus, there are totally 18 directional sub-bands and one approximate sub-band from coarse to fine scales for the real or imaginary part. The corresponding sub-dictionary for each sub-band is trained with the sub-band coefficient of initial image $\boldsymbol{x}$, in which fitting errors of all selected training patches are taken into account using the corresponding sub-dictionaries. Note that data patches of training sub-dictionaries in the MS UDCT DL process are not the image blocks themselves, but a fraction of the UDCT decomposition coefficient patches in each sub-band of each scale for the initial image. This allows further reduction of the correlation between the UDCT coefficients, giving rise to representing the image more sparsely. All sub-dictionaries are learned in the same way. Thus, a multi-scale hierarchical dictionary is achieved to permit a sparser representation for CS-MRI reconstruction.

Each sub-dictionary is trained by solving the following optimization formulation:

$$\begin{aligned}&\min_{\boldsymbol{D},\Gamma}\sum_{b=1}^{B}\sum_{i,j}\|\boldsymbol{R}_{ij}(\boldsymbol{\Psi x})_b-\boldsymbol{D}_b\boldsymbol{\alpha}_{ij}\|_2^2\\&\text{s.t.}\quad\|\boldsymbol{\alpha}_{ij}\|_0\le T_b,\quad\forall i,j.\end{aligned}\tag{7}$$

This problem describes each maximum overlapped patch for each UDCT coefficient sub-band $\boldsymbol{R}_{ij}(\boldsymbol{\Psi x})_b$ as the sparsest representation $\boldsymbol{\alpha}_{ij}$ over the unknown sub-dictionary $\boldsymbol{D}_b$, and aims to find the proper representations and the sub-dictionary. It constrains the sparsity of each UDCT coefficient sub-band by the sparsity constraint T_b and obtains the best fit for data. All sub-dictionaries constitute the multi-scale dictionaries. In each sub-band, randomly selected overlapped UDCT coefficient patches are used to train $\boldsymbol{D}_b$ by dealing with Eq. (7). The overlapping stride l is defined as the distance in pixels between relevant pixel locations in adjacent patches (Ravishankar and Bresler, 2011). The patches for l=1 are said to possess maximum overlap, which means every pixel (i, j) in each sub-band (except the ones near the right and bottom boundaries) would be the top-left corner of a square 2D patch. Each sub-dictionary $\boldsymbol{D}_b$ is learned from the corresponding UDCT sub-band coefficient patches $(\boldsymbol{\Psi x})_b$ by K-SVD (Aharon *et al*., 2006) with both fixed error threshold and sparsity threshold constraints (Ravishankar and Bresler, 2011), which costs less time than general K-SVD in solving the DL problem in Eq. (7). The proposed MS UDCT DL is described in Algorithm 1.

Algorithm 1 Multi-scale UDCT dictionary learning (MS UDCT DL) for solving sub-problem (7)

Require: the sample data applied to train multi-scale dictionary $\boldsymbol{x}$, the number of decomposition levels J, the number of atoms per sub-dictionary K, the size of atoms in each sub-dictionary n, UDCT operator $\boldsymbol{\Psi}$, scale indicator v_J

1: Initialize multi-scale dictionary: use the UDCT to decompose the initial image into $B=J\times 2v_J+1$ sub-bands. Set each initial sub-dictionary $\boldsymbol{D}_b\in\mathbb{C}^{n\times K}$ (b=1, 2, …, B) to the combination of left singular vectors of the corresponding UDCT coefficient sub-bands using singular value decomposition and some random columns of training coefficient matrix

2: **for** b=1, 2, …, B **do**

3: Divide $(\boldsymbol{\Psi x})_b$ to maximum overlapped patches $\sum_{ij}\widetilde{\mathbf{coef}_{ij}}$, where coefficient patch $\widetilde{\mathbf{coef}_{ij}}=\boldsymbol{R}_{ij}(\boldsymbol{\Psi x})_b$ is denoted as a column vector of size $n\times 1$

4: Arrange $\sum_{ij}\widetilde{\mathbf{coef}_{ij}}$ in a row vector sequentially

5: Extract a fraction column of $\sum_{ij}\widetilde{\mathbf{coef}_{ij}}$ randomly

6: Apply the K-SVD algorithm to the random partial columns of $\sum_{ij}\widetilde{\mathbf{coef}_{ij}}$ to train each sub-dictionary $\boldsymbol{D}_b$, separately

7: **end for**

8: **Return** set $\boldsymbol{D}$ of all sub-dictionaries $\boldsymbol{D}_b\in\mathbb{C}^{n\times K}$ (b=1, 2, …, $J\times 2v_J+1$)

The difference between patch-based DL in the image domain and the presented MS DL in the UDCT domain in Algorithm 1 is that the former emphasizes only global features between image pixels, while for the latter even a small patch of the fine decomposition level affects a large area in the image domain to permit not only the global but also the local perspective to represent MR images. Additionally, each sub-dictionary is learned from the maximum overlapped coefficient patches of the sub-band in the UDCT domain, to create a 'richness' in the training sub-band that generates a level of shift invariance in the resulting sub-dictionary. The obtained multi-scale dictionary coming in the form of the collection of all sub-dictionaries is considered to represent an image more sparsely and less computationally complex because representations of an image in terms of the UDCT are decomposed into a series of relatively sparse coefficient sub-bands of various scales, directions, and sizes.

3.2 PB C-SALSA for MR image reconstruction

In this section, the UDCT DL-based reconstruction method for CS-MRI is presented, making use of trained dictionaries for the UDCT coefficient sub-bands in the first stage. The ill-posed linear inverse problem of estimating $\hat{\boldsymbol{x}}$ from $\boldsymbol{y}$ in CS-MRI requires regularization or prior information. Corresponding to the aforementioned sparsity structure, we extend the constraint splitting augmented Lagrangian shrinkage algorithm (C-SALSA) (Afonso *et al*., 2011) to patch-based C-SALSA (PB C-SALSA) since it possesses some wonderful properties such as high generality, efficiency, and stability. Then a new balanced sparse model in tight frame is proposed for CS-MRI (Liu *et al*., 2015), and an efficient numerical algorithm to solve the optimization problem (C-SALSA-B) is exploited to achieve faster convergence.

The framework to obtain a numerical solution for the UDPC is shown in Fig. 3. We propose the following model to reconstruct $\hat{\boldsymbol{x}}$ by solving a constrained optimization problem:

$$\begin{aligned}\widetilde{\boldsymbol{\Gamma}}_{\mathrm{r}}&=\arg\min_{\boldsymbol{\Gamma}_{\mathrm{r}}}\phi(\boldsymbol{\Gamma}_{\mathrm{r}})=\arg\min_{\boldsymbol{x}}\phi((\boldsymbol{D}_b)^{\#}(\boldsymbol{\Psi x})_b)\\ &\text{s.t. }\left\|\boldsymbol{F}_{\mathrm{u}}\boldsymbol{x}-\boldsymbol{y}\right\|_2\le\varepsilon,\end{aligned}\tag{8}$$

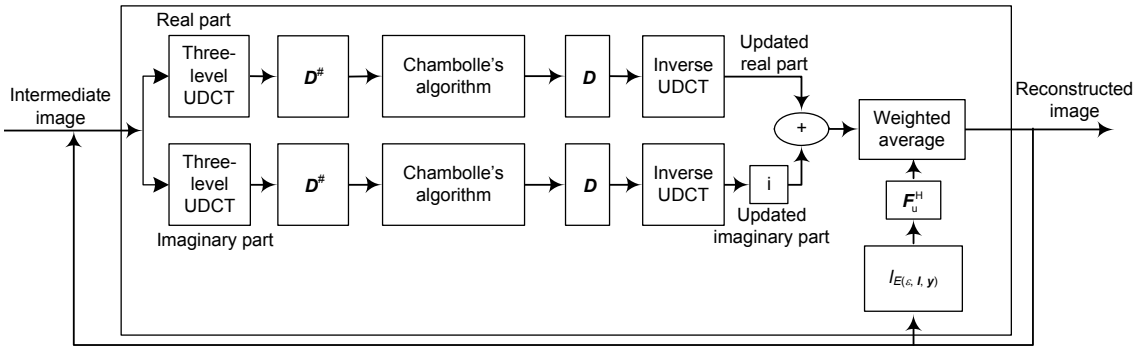


Fig. 3 Patch-based constraint splitting augmented Lagrangian shrinkage algorithm reconstruction

where $\boldsymbol{D}_b$ denotes the trained sub-dictionary according to Eq. (7) in the UDCT domain, $\boldsymbol{\Gamma}_\mathrm{r}$ the representation coefficients of the UDCT sub-band coefficients $(\boldsymbol{\Psi x})_b$ over the corresponding sub-dictionary $\boldsymbol{D}_b$, i.e., $\boldsymbol{\Gamma}_\mathrm{r}=(\boldsymbol{D}_b)^{\#}(\boldsymbol{\Psi x})_b$, ϕ the regularization function with respect to the representation coefficients of UDCT coefficients over the corresponding dictionary. Eq. (8) involves minimizing a convex, nonsmooth regularization function, under the constraint that the solution explains the observed k-space data sufficiently well. This equation updates $\boldsymbol{\Gamma}_\mathrm{r}$ when the trained dictionaries are fixed through the optimization problem.

Then the UDCT sub-band coefficients are deduced by the representation coefficients and the corresponding sub-dictionaries. The reconstructed image can be obtained after the inverse transform is applied. That is, $\hat{\boldsymbol{x}}=\boldsymbol{\psi}^{\mathrm{H}}\left(\sum_{b=1}^{B}\boldsymbol{D}_b\widetilde{\boldsymbol{\Gamma}}_\mathrm{r}\right)$, where $\boldsymbol{\psi}^{\mathrm{H}}$ denotes the inverse UDCT operator.

PB C-SALSA for solving optimization problems of form (8) is proposed to be derived from the modification of C-SALSA corresponding to the dictionary learning structure in the UDCT domain. By adding the indicator function of the feasible set, the ellipsoid $E(\varepsilon, \boldsymbol{F}_\mathrm{u}, \boldsymbol{y})=\{\boldsymbol{x}\mid \|\boldsymbol{F}_\mathrm{u}\boldsymbol{x}-\boldsymbol{y}\|_2\leq\varepsilon\}$, to objective function (8), the original constrained problem (8) turns out to be the following unconstrained one:

$$\min_{\boldsymbol{x}}\ \lambda_1\sum_{b=1}^{B}\phi((\boldsymbol{D}_b)^{\#}(\boldsymbol{\Psi x})_b)+\lambda_2 l_{E(\varepsilon,\boldsymbol{I},\boldsymbol{y})}(\boldsymbol{F}_\mathrm{u}\boldsymbol{x}), \quad (9)$$

where parameter $\lambda_1\geq 0$ measures the weight of the regularization term, $\lambda_2\geq 0$ measures the weight of k-space data fidelity, and $E(\varepsilon, \boldsymbol{I}, \boldsymbol{y})$ is simply a closed ε-radius Euclidean ball centered at $\boldsymbol{y}$. The definition of the Moreau proximal mapping associated with the regularizer ϕ (Combettes and Wajs, 2005) and the indicator function $l_{E(\varepsilon, \boldsymbol{I}, \boldsymbol{y})}$ can be found in Afonso *et al*. (2011). This mapping can be efficiently computed and contributes to the universality of C-SALSA for any kind of convex regularized term including wavelet-based (with orthogonal or frame-based representations) or TV regularization. If $\phi(\cdot)\equiv\|\cdot\|_1$, then $\Theta_{\lambda\phi}$ is simply an element-wise soft threshold. If ϕ is the TV-norm, one of several available iterative algorithms (Chambolle, 2004; Dahl *et al*., 2010) is used to compute the corresponding function $\Theta_{\lambda\phi}$. In our implementation of PB C-SALSA, the powerful TV is chosen as the regularizer due to its property of preserving sharp edges. Then the corresponding Moreau proximal mapping can be computed using Chambolle's algorithm (Chambolle, 2004).

The resulting unconstrained problem (9) is then transformed into a different constrained problem by the application of a variable splitting (VS) operation by defining $\boldsymbol{u}_0=\{(\boldsymbol{D}_b)^{\#}(\boldsymbol{\Psi x})_b|b=1, 2, \ldots, B\}$ and $\boldsymbol{v}_0=\boldsymbol{F}_\mathrm{u}\boldsymbol{x}_0$. All nonoverlapping vector-form patches (size $n\times 1$) are arrayed to produce a matrix from the UDCT sub-band of the intermediate image. Results of $(\boldsymbol{D}_b)^{\#}$ multiplying with the above matrix are the representation coefficients $\boldsymbol{\Gamma}_\mathrm{r}$ of the UDCT sub-band coefficients over the dictionary. They are generally not sparse but easier to handle in our reconstruction approach. The proximal operator may be regarded as the generalized projection operator. So, the Moreau proximal map of $l_{E(\varepsilon,\boldsymbol{I},\boldsymbol{y})}$ is simply the orthogonal projection of $\boldsymbol{F}_\mathrm{u}\boldsymbol{x}$ on the closed ε-radius ball centered at $\boldsymbol{y}$, which can be attacked by

$$\boldsymbol{\Theta}_{l_{E(\varepsilon,\boldsymbol{I},\boldsymbol{y})}}(\boldsymbol{v})=\boldsymbol{y}+\begin{cases}\varepsilon\dfrac{\boldsymbol{v}-\boldsymbol{y}}{\|\boldsymbol{v}-\boldsymbol{y}\|_2}, & \|\boldsymbol{v}-\boldsymbol{y}\|_2>\varepsilon,\\ \boldsymbol{v}-\boldsymbol{y}, & \|\boldsymbol{v}-\boldsymbol{y}\|_2\leq\varepsilon.\end{cases} \quad (10)$$

Then the alternating direction method of multipliers (ADMM) (Gabay and Mercier, 1976; Eckstein and Bertsekas, 1992) is introduced to solve two sub-problems in Eq. (9) in terms of representation coefficients $\boldsymbol{u}_k$ and k-space measurements $\boldsymbol{v}_k$ separately, in which k is the index of C-SALSA iterations. After applying C-SALSA, we arrive at the UDPC reconstruction algorithm described in Algorithm 2.

Algorithm 2 MS UDCT DL based MR image reconstruction by PB C-SALSA (UDPC) for solving problem (8)

Require: k-space data $\boldsymbol{y}$, the undersampled Fourier operator $\boldsymbol{F}_\mathrm{u}$, and UDCT operator $\boldsymbol{\Psi}$

1: Set $\boldsymbol{x}_0=\boldsymbol{F}_\mathrm{u}^{\mathrm{H}}\boldsymbol{y}$, $k=0$, and choose $\lambda>0$, $J=3$, $2v_J=6$, $B=J\times 2v_J+1$, $r=1$ or 2 ($r=1$, 2 denotes the real and imaginary parts of an image, respectively). The initial representation coefficient of the UDCT sub-band over sub-dictionary $\boldsymbol{D}_b$ is denoted as $\boldsymbol{u}_0^{(r)}$:

$$\boldsymbol{u}_0^{(r)}=\left\{\boldsymbol{u}_{0b}^{(r)}\right\}=\left\{(\boldsymbol{D}_b^{(r)})^{\#}(\boldsymbol{\Psi x}_0^{(r)})_b\mid b=1,2,\cdots,B\right\},$$

$$\boldsymbol{v}_0=\boldsymbol{y},\ \Delta\boldsymbol{u}_0^{(r)}=\boldsymbol{0}\times\boldsymbol{u}_0^{(r)},\ \Delta\boldsymbol{v}_0=\boldsymbol{0}\times\boldsymbol{v}_0,$$

where $\boldsymbol{u}_{0b}^{(r)}$ denotes the coefficients in terms of sub-dictionary $\boldsymbol{D}_b$, $(\boldsymbol{\Psi x}_0^{(r)})_b$ denotes the sub-band coefficients of $\boldsymbol{\Psi x}_0^{(r)}$, and the set of all sub-dictionaries is

$$\boldsymbol{D}^{(r)}=\left\{\boldsymbol{D}_b^{(r)}\mid b=1,2,\cdots,B,\ r=1,2\right\}$$

2: **Repeat**

3: $\boldsymbol{s}_k=\boldsymbol{\Psi}^{\mathrm{H}}\boldsymbol{D}^{(1)}(\boldsymbol{u}_k^{(1)}+\Delta\boldsymbol{u}_k^{(1)})+1\mathrm{i}\times\boldsymbol{\Psi}^{\mathrm{H}}\boldsymbol{D}^{(2)}(\boldsymbol{u}_k^{(2)}+\Delta\boldsymbol{u}_k^{(2)})+\boldsymbol{F}_{\mathrm{u}}^{\mathrm{H}}(\boldsymbol{v}_k+\Delta\boldsymbol{v}_k)$

4: $\boldsymbol{x}_{k+1}=\left[\sum_{r=1}^{2}(\boldsymbol{\Psi}^{\mathrm{H}}\boldsymbol{D}^{(r)})(\boldsymbol{D}^{(r)})^{\#}(\boldsymbol{\Psi x})+\boldsymbol{F}_{\mathrm{u}}^{\mathrm{H}}\boldsymbol{F}_{\mathrm{u}}\right]^{-1}\boldsymbol{s}_k$

5: **for** b=1, 2, …, B **do**

6: **for** r=1, 2 **do**

7: Moreau proximal mapping function of $\boldsymbol{u}_b^{(r)}$:

$\boldsymbol{u}_{(k+1)b}^{(r)}=\boldsymbol{\Theta}_{\lambda\phi}[(\boldsymbol{D}_b^{(r)})^{\#}(\boldsymbol{\Psi x}_{k+1}^{(r)})_b-\Delta\boldsymbol{u}_{kb}^{(r)}]$

8: $\Delta\boldsymbol{u}_{(k+1)b}^{(r)}=\Delta\boldsymbol{u}_{kb}^{(r)}-(\boldsymbol{D}_b^{(r)})^{\#}(\boldsymbol{\Psi x}_{k+1}^{(r)})_b+\boldsymbol{u}_{(k+1)b}^{(r)}$

9: **end for**

10: **end for**

11: Assemble all $\boldsymbol{u}_{(k+1)b}^{(r)}$ to form the updated $\boldsymbol{u}_{(k+1)}^{(r)}$, and assemble all $\Delta\boldsymbol{u}_{(k+1)b}^{(r)}$ to form the updated $\Delta\boldsymbol{u}_{(k+1)}^{(r)}$

12: Moreau proximal mapping function of $\boldsymbol{v}$:

$$\boldsymbol{v}_{k+1}=\boldsymbol{\Theta}_{l_{E(\varepsilon,\boldsymbol{I},\boldsymbol{y})}}(\boldsymbol{F}_{\mathrm{u}}\boldsymbol{x}_{k+1}-\Delta\boldsymbol{v}_k)$$

13: $\Delta\boldsymbol{v}_{k+1}=\Delta\boldsymbol{v}_k-\boldsymbol{F}_{\mathrm{u}}\boldsymbol{x}_{k+1}+\boldsymbol{v}_{k+1}$

14: $k\leftarrow k+1$

15: **Until** some stopping criterion is satisfied

16: **Return** $\hat{\boldsymbol{x}}$

Then the image in the spatial domain is acquired as the result of the inverse UDCT to the UDCT coefficients. The sub-problem with respect to $l_{E(\varepsilon,\boldsymbol{I},\boldsymbol{y})}$ can be efficiently solved via Eq. (10). The ultimate reconstructed image $\hat{\boldsymbol{x}}$ is the result of the weighted average between the above regularization penalty and the result of Eq. (10) for one iteration. The flowchart for UDPC reconstruction in Fig. 3 shows clearly the reconstruction process based on the proposed sparsifying model.

The two stages of the UDPC are implemented alternatively. The previous reconstructed image serves as the initial image of the training dictionary in the next iteration, again to guarantee new features of the image available in the current dictionary, resulting in a superior reconstruction performance. This procedure is executed in a loop until some stopping criterion is satisfied. The specified number of iterations or the difference of the mean-squared error between two consecutive iterations is assigned to the stopping criterion for PB C-SALSA reconstruction. PB C-SALSA fitting multi-scale dictionary structure can achieve fast convergence under the guarantee for the convergence of C-SALSA.

3.3 Summary for UDPC

Natural images generally exhibit meaningful structures and features over many scales, and can be analyzed and described efficiently by multi-scale constructions. Considering these reasons, a novel sparse model by training dictionaries with multi-scale structure in the UDCT domain is explored in our proposed UDPC-based MRI reconstruction from undersampled *k*-space data. This sparsity method maintains the advantages of both redundant sparse representation and MGA, thereby matching significant constituent structures of the image in multiple scales and multiple directions with a small number of nonzero coefficients, in contrast to multi-scale analysis alone or single-scale dictionary learning alone. The modified PB C-SALSA fitting multi-scale dictionary structure for reconstruction is presented in our work, which leads to a stable and rapid convergence with high-quality reconstruction. It means that the value of reconstruction parameter needs only a general estimation for most images, manifesting the flexibility and robustness of our algorithm. Therefore, our proposed multi-scale UDCT dictionary can finely match the significant features of images with the ability of superior sparse representation for MR image reconstruction under high undersampled factors. The details and fine features from different scales in MR images can be preserved in CS-MRI reconstruction with a small number of coefficients. In this way, UDPC outperforms the existing methods based on predefined transforms alone and trained dictionaries alone. In addition to all advantages of C-SALSA, the proposed UDPC is applicable to complex-valued data, which promotes CS-MRI theory much closer to the practical MRI application, along with rapid convergence and ease of implementation, giving rise to preeminent performance in CS-MRI.

3.4 Computational cost

As described in Algorithm 2, the UDPC contains alternating operations in the UDCT domain and *k*-space. The complexity of the UDPC algorithm is dominated by the MS UDCT DL stage, which

consists of 2D UDCT decomposition and each sub-dictionary training, and the related sparse coding steps. The complexity of 2D UDCT is equal to the redundancy ratio multiplied with the complexity of FFT on data ($O(P\log P)$) (Nguyen and Chauris, 2010). The estimation of the redundancy ratio in Nguyen and Chauris (2010) shows that 2D UDCT has an acceptable redundancy ratio of less than 4. The training of the dictionary and sparse coding of patches are performed using the efficient implementations of a K-SVD with both sparsity level and error threshold (Ravishankar and Bresler, 2011), which works for both real- and complex-valued training signals. With the patch size denoted by n, the number of all overlapping patches by $N=P/l^2$, the number of patches used for training by L, the number of dictionary atoms by K, and the number of iterations in learning by J, the computation in MSDL is dominated by the cost of sparse-coding, which scales as $O(LNKnT_bJ)$. Given overlap stride $l=1$ and the number of patches $N=P$, we have $O(PKnT_bJ)>>O(P\log P)$. This indicates that the sparse coding steps in Eq. (7) dominate the computational cost. As analyzed in Afonso *et al*. (2011), the computational complexity of PB C-SALSA for solving Eq. (8) in the proposed UDPC is dominated by the required sparse matrix-vector multiplications, vector additions, and the computation of the Moreau proximal map, which have a cost of $O(P\log P)$ for a wide choice of regularizers and frame representations. It is not dominated compared with the sparse coding step in the MS UDCT DL stage. In summary, the computational cost of the proposed UDPC algorithm requires approximately $O(LNKnT_bJ)$ operations. Increasing the overlap stride l can reduce the number of overlapping patches, resulting in the reduction of complexity.

4 Experiments

4.1 Experimental setup

To evaluate the performance of the proposed approach, we performed a large number of experiments on in vivo MR scan and standard phantoms, which contain a representative T2-weighted axial MR complex image of human brain and water phantom complex data (from the Computational Imaging Group at Xiamen University) (Qu *et al*., 2012; 2014; Ning *et al*., 2013), together with a realistic phantom for CS-MRI reconstruction (simulated data). To test the performance of the UDPC with different sampling patterns, 2D variable-density random sampling pattern (Trzasko and Manduca, 2009), Cartesian sampling with random phase encode scheme, and the pseudo radial line sampling were used to undersample the k-space data. Densities of MR images were normalized to [0, 1]. In this section, experiments at a variety of undersampling factors and sampling schemes were performed. The performance of the proposed UDPC was compared with those of six related methods: the leading LDP method (Lustig *et al*., 2007), adopting four levels of Daubechies wavelet decomposition for sparsifying image and C-SALSA (Afonso *et al*., 2011) (wav_CSALSA for short), DLMRI (Ravishankar and Bresler, 2011), PBDW (Qu *et al*., 2012), PBDWS (Ning *et al*., 2013), and PANO (Qu *et al*., 2014) reconstruction.

All experiments were implemented in MATLAB R2011b. Computations were performed employing a 64-bit Windows 7 operating system workstation with an Intel Xeon E5 CPU at 2.80 GHz and 8 GB memory.

Parameters needed in our algorithm were set as follows. The UDCT was of three levels, with each level possessing six directions acting on the initial image, thus obtaining 19 coefficient sub-bands including an approximation sub-band and six high-pass directional sub-bands of each level. We used the UDCT coefficient patches of size $n=8\times8$ to train overcomplete sub-dictionaries of $K=80$ atoms using $J=20$ iterations in learning of K-SVD. These parameters were chosen based on empirical tradeoffs between performance and efficiency. Maximum overlap stride $l=1$ for adjacent patches was applied to divide each sub-band into maximum overlapped patches for training each sub-dictionary. The MSDL stage employed 20 iterations, and fraction of patches $L=16\,000$ for training. The fixed sparsity thresholds of $T_{b1}\approx0.1\times n$ for approximation sub-band and $T_{b2}\approx(0.02–0.05)\times n$ for high-pass directional sub-bands were employed to train different sub-dictionaries. In the process of training sub-dictionaries, the coefficient matrix was split into 8×8 patches. Data of each patch were arranged to a column vector. All patches formed a new matrix in the form of a column vector. Then left singular vectors obtained by SVD for the new matrix

constituted a part of the initial sub-dictionary in the K-SVD learning stage. The other part consisted of some random columns of the new coefficient matrix. Each directional sub-dictionary of each scale was trained independently using the K-SVD algorithm. Each of the maximum overlapped patches was sparse-coded subject to the corresponding sparse threshold T_b. In the patch-based C-SALSA reconstruction process, the UDCT coefficient sub-bands were split into maximum overlapping patches (Ravishankar and Bresler, 2011). To ensure convergence, PB C-SALSA performed 50 inner iterations when the trained multi-scale dictionary $\{\boldsymbol{D}_b\}$ was fixed. The whole alternating procedure between MSDL and PB C-SALSA reconstruction in UDPC needed only to perform a relatively small number of iterations, such as one or two outer iterations, under a variety of sampling trajectories and acceleration factors.

The numerical metrics of quality assessment for reconstructed images in this study are as follows:

1. Peak signal-to-noise ratio (PSNR)

Suppose $\tilde{\boldsymbol{x}}$ is the fully sampled MR image, and $\hat{\boldsymbol{x}}$ the reconstructed image. Then PSNR is defined as

$$\mathrm{PSNR}=20\lg\left(255/\sqrt{\mathrm{MSE}}\right)(\mathrm{dB}), \tag{11}$$

where $\mathrm{MSE}=\dfrac{1}{M\times N}\sum_{i=0}^{M-1}\sum_{j=0}^{N-1}(\tilde{x}(i,j)-\hat{x}(i,j))^2$ ((M, N) represents the size of the MR image).

2. Structural similarity (SSIM) (Wang *et al*., 2004)

SSIM represents the structural similarity index based on the degradation of structural information. The detailed explanation can be found in Wang *et al*. (2004).

3. Transferred edge information (TEI) (Qu *et al*., 2002)

TEI measures the amount of transferred edge information, which is defined as

$$\mathrm{TEI}=Q_g^{\tilde{x}\hat{x}}Q_a^{\tilde{x}\hat{x}}, \tag{12}$$

where $Q_g^{\tilde{x}\hat{x}}$ and $Q_a^{\tilde{x}\hat{x}}$ represent the edge strength and orientation preservation value, respectively (the meanings of subscripts *a* and *g* are referred to Xydeas and Petrović, 2000).

4. Relative l_2-norm error (RLNE) (Qu *et al*., 2010)

To evaluate the reconstruction error, we use RLNE, which is defined as

$$e(\hat{\boldsymbol{x}})=\frac{\|\hat{\boldsymbol{x}}-\tilde{\boldsymbol{x}}\|_2}{\|\tilde{\boldsymbol{x}}\|_2}. \tag{13}$$

5. Concordance correlation coefficient (CCC) (Lin, 1989)

We used Lin's CCC (Lin, 1989) to measure the agreement between the reconstructed image ($\hat{\boldsymbol{x}}$) and the full sampling result ($\tilde{\boldsymbol{x}}$). The degree of concordance between the two measures can be characterized by the expected value of their squared difference:

$$E[(\hat{\boldsymbol{x}}-\tilde{\boldsymbol{x}})^2]=(\mu_{\hat{x}}-\mu_{\tilde{x}})^2+\sigma_{\hat{x}}^2+\sigma_{\tilde{x}}^2-2\rho\sigma_{\hat{x}}\sigma_{\tilde{x}}, \tag{14}$$

where μ denotes the mean, σ denotes the variance, and ρ is the Pearson correlation coefficient. The formula of CCC is as follows:

$$\begin{aligned}\mathrm{CCC}&=1-\frac{E[(\hat{\boldsymbol{x}}-\tilde{\boldsymbol{x}})^2]}{E[(\hat{\boldsymbol{x}}-\tilde{\boldsymbol{x}})^2\mid\rho=0]}\\&=\rho\frac{2}{\dfrac{(\mu_{\hat{x}}-\mu_{\tilde{x}})^2}{\sigma_{\hat{x}}\sigma_{\tilde{x}}}+\dfrac{\sigma_{\hat{x}}}{\sigma_{\tilde{x}}}+\dfrac{\sigma_{\tilde{x}}}{\sigma_{\hat{x}}}}\\&=\rho\frac{2}{\upsilon^2+\omega+1/\omega}=\rho\chi_a,\end{aligned} \tag{15}$$

where $\chi_a=\dfrac{2}{\upsilon^2+\omega+1/\omega}$, $\omega=\sigma_{\hat{x}}/\sigma_{\tilde{x}}$, and $\upsilon=|\mu_{\hat{x}}-\mu_{\tilde{x}}|/\sqrt{\sigma_{\hat{x}}\sigma_{\tilde{x}}}$.

The quantity ω is the scale shift, and the quantity υ is the location shift relative to scale. The correlation ρ is a measure of precision and the quantity χ_a is a measure of accuracy.

4.2 Performance without noise

4.2.1 Water phantom

The case of standard water phantom complex data (Qu *et al*., 2012; 2014; Ning *et al*., 2013) using the UDPC reconstruction is as shown in Fig. 4. The water phantom data were acquired on a 7 T Varian MRI system (Varian, Palo Alto, CA, USA) with the

spin echo sequence (TR/TE=2000/100 ms, 80 mm ×80 mm field of view, 2 mm slice thickness). The phantom for the MRI system was used for evaluating the image quality and testing the MRI machine performance. The phantom was well suited to assess the spatial resolution qualitatively and quantitatively over a wide range. It contained simple geometric objects that were the simple representations of anatomical structures. These objects possessed different gray scales, contrast, and spatial resolutions. They were important in evaluating MR image quality. Fig. 4a shows a fully sampled water phantom image (size 256×256). Pseudo-radial sampling pattern with sampling rate 25.97% shown in Fig. 4b was employed for the water phantom.

From the experimental results, the proposed method achieves the best resolution among all the methods (Fig. 4). The results using wav_CSALSA (Fig. 4d) and PBDW (Fig. 4f) demonstrate large errors in the smooth region. The reconstruction with LDP (Fig. 4c) and DLMRI (Fig. 4e) can suppress the artifacts, but the edges of water phantom and three circular objects show the presence of loss of considerable details and blurring around edges, whereas PBDWS, PANO, and UDPC reconstructed images exhibit clear edge details and few artifacts. The local regions of reconstructed images in Figs. 4c–4i are scaled up (by a factor of 2) for detailed observations. These details clearly show that UDPC performs slightly better in reconstructing clear curve-like details. Most image features are obtained using UDPC, which shows the superior performance of our algorithm in reconstructing the curve-like characteristics of MR images. Meanwhile, the edge of objects is much clearer and sharper than those obtained using the compared approaches. The homogeneity and image intensities of UDPC-based reconstruction appear well, which permits the visualization of small details with less graininess.

The magnitudes of the reconstruction error between the fully sampled MR image (Fig. 4a) and reconstruction results using these methods (Figs. 4c–4i) are shown on the same scale of [0, 0.25] (Figs. 4c1–4i1). We can observe that the magnitude of the reconstruction error for the proposed approach is the lowest among these methods. The error magnitudes of several compared methods (Figs. 4c1–4f1) show many more regions of high error, indicating loss of fine structured features. Their RLNEs are 0.05, 0.29, 0.07, 0.05, 0.04, 0.04, and 0.04, respectively. The assessment indices of PBDWS, PANO, and UDPC are very close. Their differences appear in the third decimal place. In other words, the proposed UDPC is highly effective at preserving more details and producing an artifact-free reconstruction (Fig. 4i), which looks close to the fully sampled image. The proposed method achieves lower reconstruction error and higher visual quality than conventional CS-MRI methods.

4.2.2 CS phantom

The CS phantom (Fig. 5a) was provided by David Smith (http://www.mathworks.com) and is a test phantom tailored to CS-MRI algorithm development. This more accurately simulates what occurs when a physical phantom is sampled in an MRI scanner. Various shaped objects with different gray scales and directional features are contained in the phantom. It is

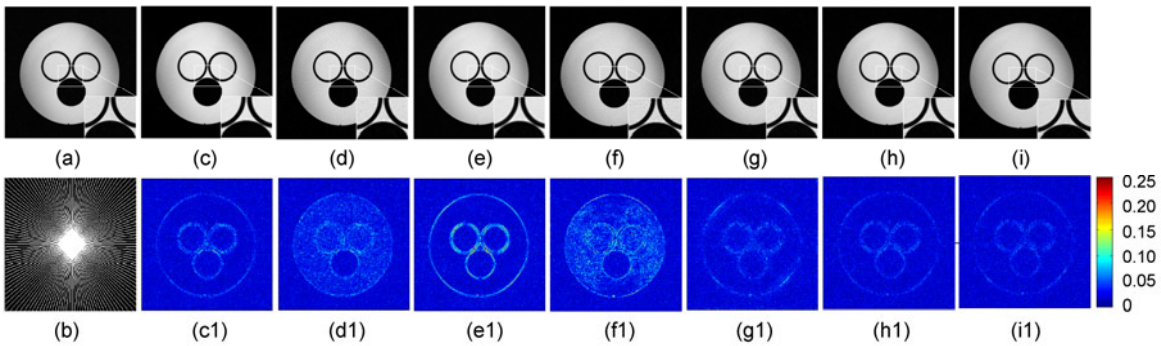


Fig. 4 Comparison of reconstruction results for a water phantom

(a) Fully sampled water phantom (256×256); (b) Simulated *k*-space trajectory (pseudo-radial lines mask with 25.97% sampling rate); (c)–(i) Reconstruction using LDP, wav_CSALSA, DLMRI, PBDW, PBDWS, PANO, and the proposed UDPC, respectively; (c1)–(i1) Magnitudes of reconstruction errors for (c)–(i) with the scale of [0, 0.25], respectively

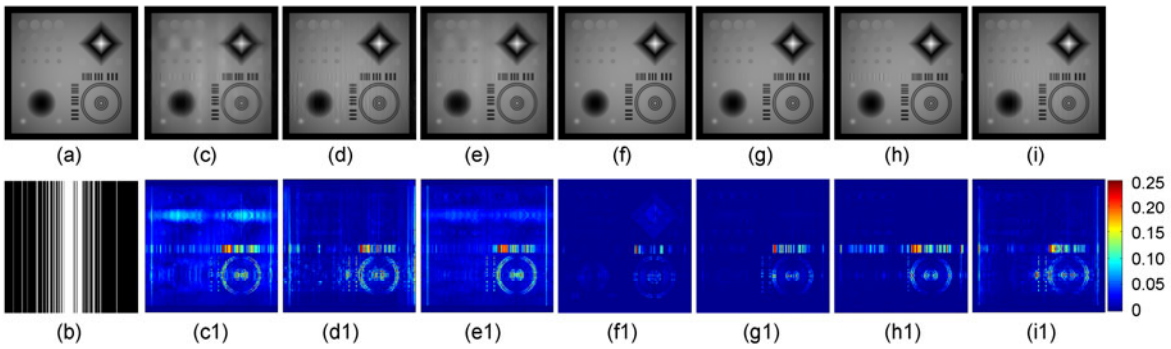


Fig. 5 Comparison of reconstructions using different methods for CS phantom

(a) Reference CS phantom image (256×256); (b) Cartesian sampling mask with sampling rate 25%; (c)–(i) Reconstruction using LDP, wav_CSALSA, DLMRI, PBDW, PBDWS, PANO, and UDPC, respectively; (c1)–(i1) Magnitudes of reconstruction errors for (c)–(i) with the scale of [0,0.25], respectively

designed to be non-sparse under a gradient transform and to contain features difficult to reproduce with partial Fourier sampling. This phantom can be used to evaluate the quality and accuracy of CS-MRI reconstruction algorithms in the noise-free domain so that real-world applications of CS-MRI may be improved. CS data acquisition is simulated by undersampling the 2D discrete Fourier transform of the CS phantom image. A variable-density Cartesian sampling pattern with a sampling rate of 25% is employed on CS phantom data (Fig. 5b). Gibbs ringing artifacts are prevalent at the object boundaries in the reconstructed images using other three methods (Figs. 5c–5e). LDP, wav_CSALSA, and DLMRI reconstructions with less contrast and loss of considerable details still show the presence of a considerable amount of aliasing and blurring around the edges. By contrast, the proposed method significantly suppresses these artifacts and preserves the edges (Fig. 5i), which shows the superiority in sharpness of small features and the level of Gibbs ringing and aliasing. More details are visible because this image has higher spatial resolution. In particular, some rings in the CS phantom have different gray scales. It is not hard to see that several rings in the middle with lower gray scales and contrast can be clearly distinguished only in the image reconstructed using our method. This is very important for clinical diagnosis, because some inchoate diseased tissue with low gray levels and contrast in MRI images cannot be faithfully rendered in MRI images if using some poor reconstruction method, which could delay the early diagnosis of some diseases. Our method performs even better in reconstructing clear edges and fine features, and can obtain preeminent reconstruction performance among state-of-the-art methods.

4.2.3 T2-weighted image of the brain

Fig. 6 shows the obtained comparison of CS-MRI reconstruction results using UDPC versus other approaches from undersampled *k*-space data for a T2-weighted MR image of the brain (Qu *et al*., 2012; 2014; Ning *et al*., 2013) (size 256×256). The complex raw data were acquired from a healthy volunteer using a 3 T Siemens Trio Tim MRI scanner using the T2-weighted turbo spin echo sequence with sequence parameters TR/TE=6100/99 ms, 220 mm×220 mm field of view, and 3 mm slice thickness. The SENSE reconstruction with reduction factor 1 is performed to compose a full *k*-space of gold standard images. These full *k*-space data are used for emulating single-channel MRI. Variable-density random sampling pattern with 20.05% sampling rate is employed.

To test the performance of our algorithm, 2D variable density random sampling with sampling rate 20.05% (Fig. 6a), was used to undersample the *k*-space data of the T2-weighted brain image. The LDP reconstruction result suffers from significant undesirable artifacts and loss of detail features, which can be seen in our latter experiments as well. Obviously, the MRI images reconstructed using LDP, wav_CSALSA, DLMRI, and PBDW suffer considerably from less contrast and less visibility in some tissue structures, which is clearly seen to have many undesirable artifacts and much loss of features. The quality of reconstruction results obtained using

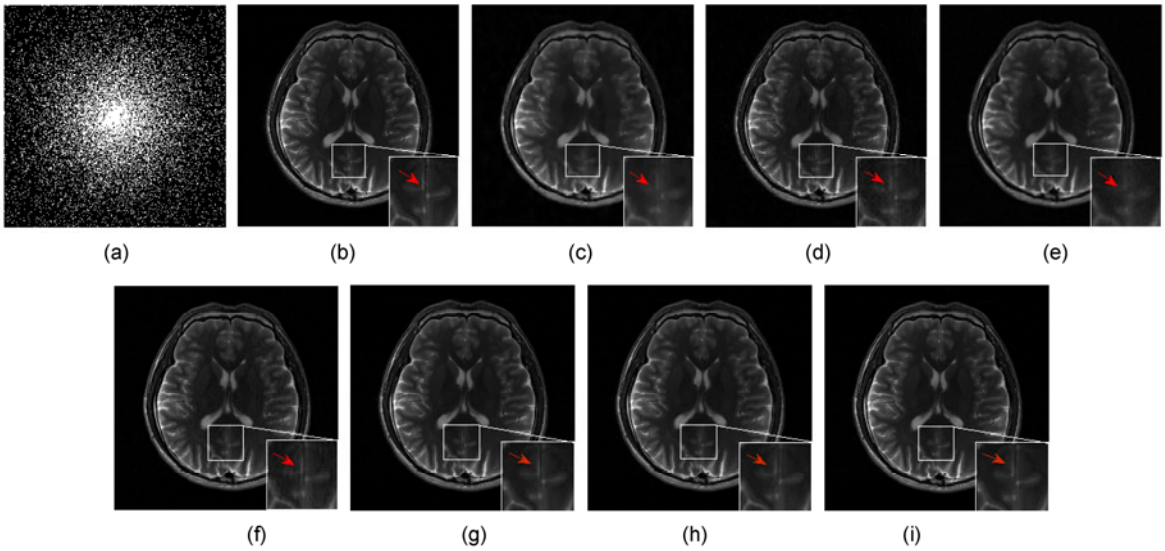


Fig. 6 Comparison of reconstruction results for the T2-weighted brain image

(a) Variable density random sampling pattern with sampling rate 20.05%; (b) Reconstructed image from fully sampled data (256×256); (c)–(i) Reconstruction using LDP, wav_CSALSA, DLMRI, PBDW, PBDWS, PANO, and the proposed UDPC, respectively. References to color refer to the online version of this figure

PBDWS, PANO, and the proposed approach has a large degree of improvement. As shown in Fig. 6i, the reconstruction using UDPC is relatively devoid of aliasing artifacts. Gibbs ringing artifacts have been drastically mitigated by employing the proposed approach compared with the others due to the sparsifying model in our method. Specifically, in comparison to other methods, the images reconstructed using UDPC can provide good contrast between gray and white matters. The anatomic structure of bilateral basal ganglia is depicted well. Cerebral cisterns and sulci are present with a clear border. The MR images reconstructed using UDPC with higher contrast and spatial resolution can preserve tissue boundary integrity and texture of tissue well. To better illustrate the reconstruction performance of UDPC for some fine features of MR images, the comparison with visual appearance is given in the zooming regions of Fig. 6 (areas marked by red arrows). The PSNRs of these methods are 31.92, 33.64, 30.99, 36.68, 38.82, 37.24, and 37.96 dB, respectively. Their RLNEs are 0.16, 0.13, 0.17, 0.09, 0.08, 0.09, and 0.09, respectively. It shows that edges are better preserved and lower reconstruction errors are achieved using UDPC than using the earlier methods. The proposed method can achieve comparable performance with PANO and PBDWS.

4.3 Performance under noisy circumstances

To demonstrate the performance of the proposed method in the noisy case, zero-mean complex Gaussian white noise with standard deviation σ=10.2 is added to the k-space data. Fig. 7 shows the reconstruction results of the proposed method and other methods from k-space noisy data in Fig. 6a, employing 2D variable density random sampling with 20.05% sampling rate. Two iterations are applied in UDPC while eight iterations are executed in DLMRI. Fig. 7a is the fully sampled noisy image, and Fig. 7b shows the magnitude of the added noise. The PSNR of the fully sampled noisy image corresponding to the reference is approximately 29.92 dB. The PSNRs of the reconstructed images in Figs. 7c–7i are 30.25, 29.41, 30.25, 30.01, 31.25, 31.47, and 31.56 dB, respectively. It is found that our method achieves the highest PSNR, which is about 1.3 dB higher than those of LDP and DLMRI, and about 2.1 dB higher than that of wav_CSALSA. The reconstruction with LDP and DLMRI is unable to sufficiently remove obvious artifacts and noise. The proposed method achieves good reconstruction quality with edge preserving and less noise. The noise is greatly reduced in reconstruction (Fig. 7i), which shows the superior ability of noise suppression. The red arrows in the

zoom of Figs. 7c–7i indicate visible details. It is observed that a much smaller reconstruction error and loss of features are obtained for UDPC compared with other methods. The proposed method is shown to acquire preeminent results even under noisy cases.

Table 1 depicts vividly the comparison of objective evaluation indices for aforementioned three kinds of noise-added data according to the proposed UDPC CS-MRI reconstruction with compared methods using three different sampling patterns with different sampling rates. The PSNRs of the fully sampled noisy image for the three cases in Table 1 are 29.9198, 34.9163, and 34.1096 dB, respectively. Only the PSNR of our proposed UDPC reconstruction is closest to that of the fully sampled image, even higher for the T2-weighted brain image. The values of four typical evaluation indices indicate that in general the proposed method obtains a better reconstruction performance and significant improvements than the other methods.

Fig. 7 and Table 1 show that the proposed method can not only obtain minor reconstruction errors and good edge-preserving characteristics, but also improve the spatial detail information and preserve the structural similarity of image over all the compared methods (especially in zooming), which can also be justified by the obtained optimal values of evaluation indices (Table 1). These objective

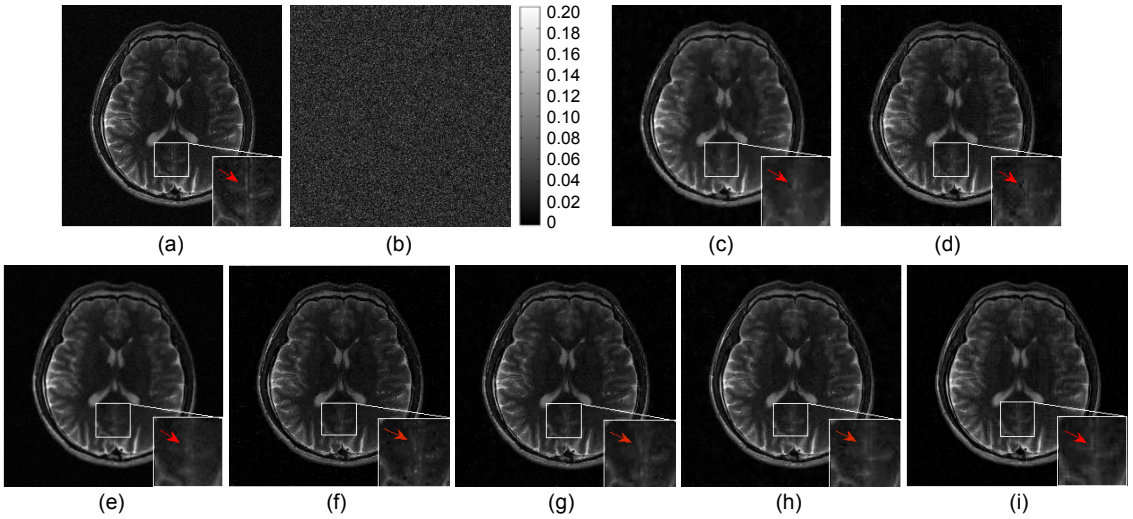


Fig. 7 Comparison of reconstruction results for noisy T2-weighted brain image data

(a) Reconstructed image from fully sampled noisy data (σ=10.2); (b) Noise magnitude in (a); (c)–(i) Reconstruction using LDP, wav_CSALSA, DLMRI, PBDW, PBDWS, PANO, and the proposed UDPC, respectively. References to color refer to the online version of this figure

Table 1 Comparison of the objective assessment indices of CS-MRI reconstruction for noise-added data via UDPC to compared methods using three different sampling patterns with different sampling rates*

Method	T2-weighted brain image & mask_VDRS_2001 & sigma=10.2				Water phantom & mask_PRS_2597 & sigma=6.5				CS phantom & mask_Carteian_25 & sigma=6.5			
	PSNR (dB)	TEI	SSIM	RLNE	PSNR (dB)	TEI	SSIM	RLNE	PSNR (dB)	TEI	SSIM	RLNE
LDP	30.25	0.52	0.79	0.19	33.57	0.68	0.72	0.04	29.04	0.57	0.88	0.09
wav_CSALSA	29.41	0.52	0.71	0.21	16.94	0.50	0.65	0.30	28.65	0.58	0.74	0.09
DLMRI	30.36	0.51	0.86	0.19	32.73	0.66	0.73	0.05	31.45	0.66	0.84	0.07
PBDW	30.01	0.57	0.77	0.19	31.43	0.66	0.70	0.06	33.33	0.68	0.87	0.05
PBDWS	31.65	0.61	0.79	0.16	34.20	0.68	0.81	0.04	33.35	0.71	0.88	0.05
PANO	31.47	0.61	0.77	0.16	34.46	0.69	0.82	0.04	32.20	0.69	0.87	0.06
UDPC	31.56	0.60	0.86	0.16	33.52	0.68	0.78	0.04	32.11	0.69	0.87	0.06

* mask_VDRS_2001 denotes variable density random sampling mask with 20.01% sampling rate, mask_ PRS_2597 denotes pseudo radial lines sampling mask with 25.97% sampling rate, and mask_Carteian_25 denotes with variable density (Cartesian sampling mask in *k*-space with 25% sampling rate)

assessment indices agree with the visual assessment.

Considering image quality assessment from various comprehensive evaluation standards, the superior reconstruction performance is achieved by the proposed UDPC CS-MRI reconstruction method. Furthermore, it has good stability in various cases.

4.4 Assessing agreement

The comparisons of Lin's CCC values (Lin, 1989) of each reconstruction method with respect to the full sampling result for phantom and in vivo reconstructions are demonstrated in Table 2. In particular, the CCC for noisy T2-weighted image of the brain is the concordance between the reconstructed image and the noisy fully sampled reconstruction image. According to the data in Table 2, the reconstruction results of the proposed method with the highest CCCs are verified.

4.5 Evaluation of the sparsity ratio

Via MSDL based on UDCT operation, a novel sparsity model is built. Thus, the multi-scale UDCT dictionary permits the optimal sparse representation of the image with intrinsic properties of the image preserved among current sparse methods. To illustrate the superiority of the proposed UDCT multi-scale dictionary in sparsifying the image, the image in Fig. 4b is used for sparse assessment. The sparsity ratios of the Daubechies wavelet, UDCT, and multi-scale UDCT dictionary in this work are compared. For Daubechies wavelet and UDCT, the sparsity ratio for the corresponding sparse transform is defined to be the proportion of coefficients whose absolute values are larger than the mean absolute value to the total number of coefficients. For the UDCT multi-scale dictionary, the sparsity ratio is designated as the ratio of nonzero elements in one coefficient set in terms of the corresponding sub-dictionary. The sparsity ratios of the three methods compared are given in Table 3. It is found that the optimal sparse percentage for the coefficient matrix is obtained by the multi-scale UDCT dictionary method. The sparsity ratio has a large influence on the convergence speed of the reconstruction algorithm. The sparsity ratio comparison illustrates the efficiency of our algorithm in representing the image with optimal sparsity, laying the foundation for reconstructing images under high undersampling factors.

4.6 Computation time

The computation time of these reconstruction methods is summarized in Table 4. The iteration stopping criterion is that the norm of the difference between the reconstructed image and the fully sampled image of two successive iterations for reconstruction is less than a certain threshold. Table 4 shows clearly that in DLMRI, the average runtime for single-scale dictionary training (20 K-SVD iterations) is 100 s or so, and for the reconstruction procedure is 131 s. The wav_CSALSA method using 85 iterations

Table 2 Comparisons of Lin's CCC values of each reconstruction method with respect to the full sampling result for phantom and in vivo reconstructions

Method	Lin's CCC value			
	Water phantom	CS phantom	T2-weighted brain image	Noisy T2-weighted brain image
LDP	0.9984	0.9849	0.9796	0.9430
wav_CSALSA	0.9120	0.9880	0.9868	0.9408
DLMRI	0.9981	0.9885	0.9746	0.9485
PBDW	0.9981	0.9888	0.9887	0.9489
PBDWS	0.9985	0.9902	0.9901	0.9516
PANO	0.9988	0.9930	0.9904	0.9525
UDPC	0.9988	0.9932	0.9904	0.9535

Table 3 Sparsity ratios of different sparse methods for four-level decompositions

Method	Sparsity ratio				
	Low pass	Fourth level	Third level	Second level	First level
Daubechies wavelet	0.4277	0.4131	0.4131	0.4131	0.4131
UDCT	0.4277	0.3506	0.3545	0.3918	0.3590
UDCT-based DL	0.0468	0.0078	0.0078	0.0154	3.91×10^{-6}

Table 4 Runtime of different reconstruction methods*

Method	LDP	wav_CSALSA	DLMRI	PBDW	PBDWS	PANO	UDPC
Runtime (s)	293.85	46.60	100.02+131.33	197.25	213.19	305.89	26.37/sub-dictionary +111

* Computations were performed on a 2.80 GHz Intel Xeon E5 workstation with 8 GB RAM for Fig. 6b with 20.05% *k*-space data. The total runtime for DLMRI and UDPC contains dictionary training time and reconstruction time

has the runtime of 46.60 s, and MS UDCT dictionary training in UDPC has a runtime of about 501 s for all 19 complex-valued sub-dictionaries and about 26.37 s for each sub-dictionary on average. The runtime of PB C-SALSA reconstruction (50 iterations) is about 111 s with complex dictionaries. The time-consuming part of the proposed method is the MS UDCT DL stage, but the computational cost of dictionary training may be decreased substantially with the use of code optimization and graphics processing units (GPU). In addition, the modified PB C-SALSA fitting multi-scale dictionary structure achieves relatively fast convergence with preeminent reconstruction performance compared with LDP and DLMRI. From the above analyses, it is obvious that our proposed UDPC method obtains the tradeoff between performance and computational cost in CS-MRI reconstruction.

5 Conclusions and future work

In this paper we proposed a novel multi-scale UDCT dictionary learning framework for CS reconstruction of MR data. The structure consists of a sparsity model of adaptive multi-scale dictionary learning in the UDCT domain together with the corresponding patch-based constraint splitting augmented Lagrangian shrinkage reconstruction algorithm. In the first stage, a complex-valued multi-scale dictionary is trained to sparsely represent complex-valued data, which helps remove aliasing and Gibbs ringing artifacts and to preserve the rich intrinsic properties of MR images. The proposed sparsity model has been proved to allow much sparser representation, resulting in significant improvement over all compared methods in CS-MRI application. In the second stage, patch-based C-SALSA enforces data fidelity with rapid convergence without degrading the reconstruction quality. Experimental results demonstrated that the proposed method can provide superior reconstruction performance at high acceleration rates with or without noise, which leads to greatly improved visual quality of the reconstruction image with much less graininess and less information loss than the others. Future work will aim at the acceleration of implementation for UDPC with the use of GPUs and the extension of proposed method to multi-coil MR technology.

Acknowledgements

The authors sincerely thank Dr. M. LUSTIG for sharing the interface framework of CS-MRI and S. RAVISHANKAR and Y. BRESLER for sharing their experimental materials and source codes. We would also like to thank Dr. Xiao-bo QU with the Computational Imaging Group at Xiamen University for sharing codes and raw *k*-space MRI data, and thank our coworkers for helpful discussions.

References

Afonso, M.V., Bioucas-Dias, J.M., Figueiredo, M.A.T., 2011. An augmented Lagrangian approach to the constrained optimization formulation of imaging inverse problems. *IEEE Trans. Image Process.*, **20**(3):681-695. [doi:10.1109/TIP.2010.2076294]

Aharon, M., Elad, M., Bruckstein, A., 2006. K-SVD: an algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Trans. Signal Process.*, **54**(11):4311-4322. [doi:10.1109/TSP.2006.881199]

Baraniuk, R., 2007. Compressive sensing. *IEEE Signal Process. Mag.*, **24**(4):118-121. [doi:10.1109/MSP.2007.4286571]

Candes, E.J., Donoho, D.L., 2004. New tight frames of curvelets and optimal representations of objects with piecewise C^2 singularities. *Commun. Pure Appl. Math.*, **57**(2):219-266. [doi:10.1002/cpa.10116]

Candes, E.J., Romberg, J., Tao, T., 2006a. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inform. Theory*, **52**(2):489-509. [doi:10.1109/TIT.2005.862083]

Candes, E.J., Romberg, J.K., Tao, T., 2006b. Stable signal rcovery from incomplete and inaccurate measurements. *Commun. Pure Appl. Math.*, **59**(8):1207-1223. [doi:10.1002/cpa.20124]

Chambolle, A., 2004. An algorithm for total variation minimization and applications. *J. Math. Imag. Vis.*, **20**(1-2):89-97. [doi:10.1023/B:JMIV.0000011325.36760.1e]

Chen, C., Huang, J., 2014. The benefit of tree sparsity in accelerated MRI. *Med. Image Anal.*, **18**(6):834-842. [doi:10.1016/j.media.2013.12.004]

Combettes, P.L., Wajs, V.R., 2005. Signal recovery by proximal forward-backward splitting. *Multiscale Model. Simul.*, **4**(4):1168-1200. [doi:10.1137/050626090]

Dahl, J., Hansen, P.C., Jensen, S.H., *et al*., 2010. Algorithms and software for total variation image reconstruction via first-order methods. *Numer. Algor.*, **53**(1):67-92. [doi:10.1007/s11075-009-9310-3]

Donoho, D.L., 2001. Sparse components of images and optimal atomic decompositions. *Constr. Approx.*, **17**(3): 353-382. [doi:10.1007/s003650010032]

Donoho, D.L., 2006. Compressed sensing. *IEEE Trans. Inform. Theory*, **52**(4):1289-1306. [doi:10.1109/TIT.2006.871582]

Eckstein, J., Bertsekas, D.P., 1992. On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators. *Math. Program.*, **55**(1): 293-318. [doi:10.1007/BF01581204]

Elad, M., 2010. Sparse and Redundant Representations: from Theory to Applications in Signal and Image Processing. Springer, New York, USA. [doi:10.1007/978-1-4419-7011-4]

Elad, M., Aharon, M., 2006. Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Trans. Image Process.*, **15**(12):3736-3745. [doi:10.1109/TIP.2006.881969]

Gabay, D., Mercier, B., 1976. A dual algorithm for the solution of nonlinear variational problems via finite element approximation. *Comput. Math. Appl.*, **2**(1):17-40. [doi:10.1016/0898-1221(76)90003-1]

Gho, S.M., Nam, Y., Zho, S.Y., *et al.*, 2010. Three dimension double inversion recovery gray matter imaging using compressed sensing. *Magn. Reson. Imag.*, **28**(10): 1395-1402. [doi:10.1016/j.mri.2010.06.029]

Huang, J., Zhang, S., Metaxas, D., 2011. Efficient MR image reconstruction for compressed MR imaging. *Med. Image Anal.*, **15**(5):670-679. [doi:10.1016/j.media.2011.06.001]

Kim, Y., Altbach, M.I., Trouard, T.P., *et al*., 2009. Compressed sensing using dual-tree complex wavelet transform. *Proc. Int. Soc. Mag. Reson. Med.*, **17**:2814.

Kim, Y., Nadar, M.S., Bilgin, A., 2012. Wavelet-based compressed sensing using a Gaussian scale mixture model. *IEEE Trans. Image Process.*, **21**(6):3102-3108. [doi:10.1109/TIP.2012.2188807]

Lewicki, M.S., Sejnowski, T.J., 2000. Learning overcomplete representations. *Neur. Comput.*, **12**(2):337-365. [doi:10.1162/089976600300015826]

Lin, L., 1989. A concordance correlation coefficient to evaluate reproducibility. *Biometrics*, **45**(1):255-268. [doi:10.2307/2532051]

Liu, Y., Cai, J., Zhan, Z., *et al.*, 2015. Balanced sparse model for tight frames in compressed sensing magnetic resonance imaging. *PLoS ONE*, **10**(4):e0119584.1-e0119584.19. [doi:10.1371/journal.pone.0119584]

Lustig, M., Donoho, D., Pauly, J.M., 2007. Sparse MRI: the application of compressed sensing for rapid MR imaging. *Magn. Reson. Med.*, **58**(6):1182-1195. [doi:10.1002/mrm.21391]

Lustig, M., Donoho, D.L., Santos, J.M., *et al*., 2008. Compressed sensing MRI. *IEEE Signal Process. Mag.*, **25**(2):72-82. [doi:10.1109/MSP.2007.914728]

Mallat, S., 2008. A Wavelet Tour of Signal Processing: the Sparse Way (3rd Ed.). Academic Press, USA.

Nguyen, T.T., Chauris, H., 2010. Uniform discrete curvelet transform. *IEEE Trans. Signal Process.*, **58**(7):3618-3634. [doi:10.1109/TSP.2010.2047666]

Ning, B., Qu, X., Guo, D., *et al*., 2013. Magnetic resonance image reconstruction using trained geometric directions in 2D redundant wavelets domain and non-convex optimization. *Magn. Reson. Imag.*, **31**(9):1611-1622. [doi:10.1016/j.mri.2013.07.010]

Ophir, B., Lustig, M., Elad, M., 2011. Multi-scale dictionary learning using wavelets. *IEEE J. Sel. Topics Signal Process.*, **5**(5):1014-1024. [doi:10.1109/JSTSP.2011.2155032]

Qu, G., Zhang, D., Yan, P., 2002. Information measure for performance of image fusion. *Electron. Lett.*, **38**(7): 313-315. [doi:10.1049/el:20020212]

Qu, X., Zhang, W., Guo, D., *et al*., 2010. Iterative thresholding compressed sensing MRI based on contourlet transform. *Inv. Probl. Sci. Eng.*, **18**(6):737-758. [doi:10.1080/17415977.2010.492509]

Qu, X., Guo, D., Ning, B., *et al*., 2012. Undersampled MRI reconstruction with patch-based directional wavelets. *Magn. Reson. Imag.*, **30**(7):964-977. [doi:10.1016/j.mri.2012.02.019]

Qu, X., Hou, Y., Lam, F., *et al*., 2014. Magnetic resonance image reconstruction from undersampled measurements using a patch-based nonlocal operator. *Med. Image Anal.*, **18**(6):843-856. [doi:10.1016/j.media.2013.09.007]

Rauhut, H., Schnass, K., Vandergheynst, P., 2008. Compressed sensing and redundant dictionaries. *IEEE Trans. Inform. Theory*, **54**(5):2210-2219. [doi:10.1109/TIT.2008.920190]

Ravishankar, S., Bresler, Y., 2011. MR image reconstruction from highly undersampled *k*-space data by dictionary learning. *IEEE Trans. Med. Imag.*, **30**(5):1028-1041. [doi:10.1109/TMI.2010.2090538]

Rubinstein, R., Zibulevsky, M., Elad, M., 2010. Double sparsity: learning sparse dictionaries for sparse signal approximation. *IEEE Trans. Signal Process.*, **58**(3): 1553-1564. [doi:10.1109/TSP.2009.2036477]

Rudin, L.I., Osher, S., Fatemi, E., 1992. Nonlinear total variation based noise removal algorithms. *Phys. D*, **60**(1-4): 259-268. [doi:10.1016/0167-2789(92)90242-F]

Trzasko, J., Manduca, A., 2009. Highly undersampled magnetic resonance image reconstruction via homotopic l_0-minimization. *IEEE Trans. Med. Imag.*, **28**(1):106-121. [doi:10.1109/TMI.2008.927346]

Wang, Z., Bovik, A.C., Sheikh, H.R., *et al*., 2004. Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Process.*, **13**(4):600-612. [doi:10.1109/TIP.2003.819861]

Xydeas, C.S., Petrović, V., 2000. Objective image fusion performance measure. *Electron. Lett.*, **36**(4):308-309. [doi:10.1049/el:20000267]

Zhu, Z., Wahid, K., Babyn, P., *et al.*, 2013. Compressed sensing-based MRI reconstruction using complex double-density dual-tree DWT. *Int. J. Biomed. Imag.*, **2013**: 907501.1-907501.12. [doi:10.1155/2013/907501]