2013 IEEE 10th International Symposium on Biomedical Imaging:
From Nano to Macro
San Francisco, CA, USA, April 7-11, 2013

UNDERSAMPLED DYNAMIC MAGNETIC RESONANCE IMAGING USING PATCH-BASED SPATIOTEMPORAL DICTIONARIES

Yanhua Wang, *Yihang Zhou*, *Leslie Ying*

Department of Biomedical Engineering, Department of Electrical Engineering
The State University of New York at Buffalo, Buffalo, NY, USA

ABSTRACT

Dynamic magnetic resonance imaging (dMRI) requires high spatial and temporal resolutions, which is challenging due to the low imaging speed. To reduce the imaging time, a patch-based spatiotemporal dictionary learning (DL) model is proposed for compressed-sensing reconstruction of dynamic images from undersampled data. Specifically, the dynamic image sequence is divided into overlapping patches along both the spatial and temporal directions. These patches are expected to be sparsely represented over a set of temporal-dependent spatiotemporal dictionaries. The images are then reconstructed from the undersampled data in (k,t) space under such sparseness constraints, where the dictionaries are learned iteratively. Alternating optimization is applied to solve the problem. Simulation results show that the proposed method is capable of preserving details in both spatial and temporal directions.

Index Terms — dynamic magnetic resonance imaging, compressed sensing, dictionary learning, spatiotemporal dictionary

1. INTRODUCTION

Dynamic magnetic resonance imaging (dMRI) plays a key role in many clinical applications, such as cardiac cine imaging, parameter mapping and functional imaging. However, high spatial-temporal resolution dMRI is still challenging due to the long scan time [1]. A lot of work (e.g., [2][3]) has been done to reduce the amount of acquired data and thus the acquisition time by exploiting the spatial and temporal correlations in the dynamic image sequence. With the emergence of compressed sensing (CS), such correlations are represented as sparseness of the image sequence under a certain sparsifying transform [4]-[7].

Conventional CS requires prior knowledge about the sparsifying transform which is essential for the recovery process. A suitable sparsifying transform should have certain structures and fast numerical implementations, such as Fourier and Wavelet transform. However, it is difficult to find a universal transform that can sparsify all signals.

The recent researches on sparse representation over learned dictionaries indicates great potential on many practical applications. Properly trained or adaptive dictionaries have the ability to offer more flexible sparse representations and provide better performances over the conventional analytical sparsifying transforms [8]-[10]. In recent years, the application of dictionary learning (DL) in MRI has attracted a lot of attentions. In [11] and [12], a dictionary is learned to sparsely represent images spatially. Although these methods can be directly extended to dMRI by reconstructing images frame by frame, the sparseness is not exploited along the temporal axis. On the other hand, the concept of DL is used for temporal sparse representation in [13][14]where the application concentrates on the parameter mapping. Most recently, a spatiotemporal DL model is proposed for dMRI with an additional low-rank constraint [15]. This method uses all the temporal data points to form the dictionary. It may not capture the local abrupt motion and thus lose some temporal details. Furthermore, the size and the number of atoms increase rapidly with the number of frames. Building such a large dictionary leads to high computational complexity due to the nonlinear operations.

In the spatially-based DL techniques, an image is divided into many overlapping patches, which are used for dictionary learning and sparse representation [11][12]. Inspired by the work in [9], we extend the patch-based DL model from the spatial domain to the spatial-temporal domain. The image sequence in dMRI is divided into overlapping patches along both spatial and temporal directions. Thus, the atoms in the corresponding dictionaries are of three dimensions. In contrast to the spatiotemporal model in [15], our model adapts to specific local spatial-temporal features. The computation time in DL is also reduced due to the smaller size of atoms. The simulation results show that the proposed method is capable of capturing local features, especially the abrupt changes along time.

This paper is structured as follows. In Section 2, the basics of DL are briefly described. Then, the proposed patch-based spatiotemporal dictionary model is introduced. Finally, the corresponding dMRI reconstruction problem is formulated and the optimization algorithm is presented.

Section 3 shows the simulation results using a cardiac cine imaging data set. The paper is concluded in Section 4.

2. THE PROPOSED METHOD

2.1. Dictionary Learning

Sparse coding over a redundant dictionary has proven to be a powerful tool in signal processing. A dictionary is a finite collection of vectors that spans a given signal space. The members of a dictionary are called atoms. A dictionary is redundant or overcomplete if the number of atoms is larger than their size. In this case, these atoms are linear dependent.

Dictionary learning focuses on building a suitable dictionary that can provide efficient sparse representations for a class of signals [10]. Given a set of signals $\mathbf{X} = \{\mathbf{x}_i\}_{i=1}^{N}$, DL aims to solve the following problem:

$$\begin{aligned} &\min_{\mathbf{D},\mathbf{A}} \ \|\mathbf{X} - \mathbf{D}\mathbf{A}\|_F^2 \\ &\text{s.t.} \ \ \|\boldsymbol{\alpha}_i\|_0 \le K, \ \ \forall i \end{aligned} \tag{1}$$

where $\|\bullet\|_F$ denotes the Frobenius norm, $\|\bullet\|_0$ is the l_0 norm, $\mathbf{D}$ stands for the dictionary to be built, $\mathbf{A}=[\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \ldots, \boldsymbol{\alpha}_N]$, $\boldsymbol{\alpha}_i$ is called the sparse representation of $\mathbf{x}_i$ over $\mathbf{D}$, and K is the sparsity constraint.

Numerous algorithms have been developed for the above DL problem and a brief review can be found in [10]. In this paper, we choose the K-SVD algorithm due to its efficiency [16]. The K-SVD algorithm consists of two stages by performing alternating optimization between $\mathbf{D}$ and $\mathbf{A}$. The first stage is called sparse approximation. In this step, $\mathbf{D}$ is fixed. The members of $\mathbf{A}$ are solved separately by using any sparse coding methods. The next stage is dictionary updating. $\mathbf{D}$ is solved column by column under a fixed $\mathbf{A}$. The singular value decomposition (SVD) operation is involved to minimize the approximation error. An initial guess of $\mathbf{D}$ is needed in the K-SVD algorithm. The over-complete discrete cosine transform (DCT) dictionary is widely used as the initial value. The combinations of other orthogonal transforms can be also used for initialization. Although the K-SVD algorithm cannot guarantee to reach a global minimum, it shows excellent performance in practical applications [8][9].

2.2. Patch-based Spatiotemporal Dictionary

To sparsify static images using DL, an image is divided into overlapping patches in order to capture local structures spatially and reduce the computation complexity at the same time [11][12]. In dynamic MRI, high coherence exhibits in temporal direction as well, which enables a sparse representation in both spatial and temporal domains. We extend the concept of patch-based learning to the temporal dimension.

Let $\mathbf{X}$ denote the dMRI image sequence and $\mathbf{X}_t$ (t=1,...,T) represent the t-th frame image. Define R_{ijk} as the operator that extracts a three dimensional patch from $\mathbf{X}$ which is located at (i, j, k) and with a predefined size of (n_f, n_p, n_t). A straightforward method is to build a single dictionary over which all the above patches can be sparsely represented. However, this approach is not suitable. The spatial structures and the temporal variations change with frames. A single dictionary cannot work well for all patches.

An appropriate solution is to use temporal dependent dictionaries. Specifically, consider a sub sequence $\mathbf{X}_{t,M}$ that contains M consecutive images, i.e. $\mathbf{X}_{t,M}$=[$\mathbf{X}_t$; $\mathbf{X}_{t+1}$;...; $\mathbf{X}_{t+M-1}$]. We divide $\mathbf{X}_{t,M}$ into overlapping patches and expect that these patches can be sparsely represented over a dictionary $\mathbf{D}_t$. The subscript t below $\mathbf{D}$ indicates that the dictionary is temporal dependent. This local dictionary model allows better adaption to both the spatial structures and temporal variations.

The corresponding DL problem can be expressed as:

$$\begin{aligned} &\min_{\mathbf{D}_t,\mathbf{A}_t} \ \sum_{i,j,k} \left\|\mathbf{R}_{ijk}\mathbf{X}_{t,M} - \mathbf{D}_t\boldsymbol{\alpha}_{ijk}\right\|_2^2 \\ &\text{s.t.} \ \ \|\boldsymbol{\alpha}_{ijk}\|_0 \le K, \ \ \forall i,j,k \end{aligned} \tag{2}$$

Here, $\mathbf{A}_t$ denotes the set containing all $\boldsymbol{\alpha}_{ijk}$. The dictionary $\mathbf{D}_t$ has a size of $P\times Q$, where $P = n_f\times n_p\times n_t$ and Q is the number of atoms. The indexes (i, j, k) cover all the available patches. Note that the condition $M \ge n_t$ is implied in this formulation. Thus, the range of k can be larger than one.

The proposed model has the ability to capture large local temporal variations. In addition, the model has a smaller atom size because n_t is much less than T. Since the redundancy factor of a dictionary is always fixed, the number of atoms is also reduced. On the other hand, as consecutive images in dMRI are similar, the dictionaries $\mathbf{D}_t$ and $\mathbf{D}_{t+1}$ are also expected to contain similar structures. Hence, the result of $\mathbf{D}_t$ can be used for the initial value when training $\mathbf{D}_{t+1}$. This can help reduce the iterations of the K-SVD algorithm in the subsequent frames [9].

2.3. Image Reconstruction

Using the above patch-based spatiotemporal DL model, we formulate the dMRI reconstruction problem as:

$$\begin{aligned} &\min_{\mathbf{X},\mathbf{D}_t,\mathbf{A}_t} \ \mathrm{TV}\left(\mathbf{X}_{t,M}\right) + \lambda_1 \left\|\mathbf{F}_t^u\mathbf{X}_{t,M} - \mathbf{Y}_{t,M}\right\|_2^2 + \\ &\qquad \lambda_2 \sum_{i,j,k} \left\|\mathbf{R}_{ijk}\mathbf{X}_{t,M} - \mathbf{D}_t\boldsymbol{\alpha}_{ijk}\right\|_2^2 \qquad t = 1,2,\cdots \\ &\text{s.t.} \ \ \|\boldsymbol{\alpha}_{ijk}\|_0 \le K, \ \ \forall i,j,k \end{aligned} \tag{3}$$

where TV($\cdot$) stands for total variation (TV) operation along the spatial dimension, $\mathbf{Y}_{t,M}$ = [$\mathbf{Y}_t$; $\mathbf{Y}_{t+1}$; … ;$\mathbf{Y}_{t+M-1}$], $\mathbf{Y}_t$ is the undersampled k-space data of the t-th frame image $\mathbf{X}_t$, $\mathbf{F}_t^u$ denotes the corresponding undersampled Fourier transforms, and λ_1 and λ_2 are tuning parameters. This formulation preserves the conventional TV regularization term to exploit the spatial sparsity, because it has proven to be effective. Further sparsity is enforced by using the spatiotemporal dictionaries. Compared with some other sparsity models in dMRI, such as **r**-f sparsity used in the k-t FOCUSS algorithm [6], it allows various sampling patterns, which offers more flexibility in practical applications.

The alternating optimization method is used to solve the above problem. The minimization procedure is divided into two steps. In one step, $\mathbf{X}_{t,M}$ is fixed. The problem becomes:

$$\begin{aligned} &\min_{\mathbf{D}_t,\mathbf{A}_t} \sum_{i,j,k} \left\| \mathbf{R}_{ijk}\mathbf{X}_{t,M} - \mathbf{D}_t \boldsymbol{\alpha}_{ijk} \right\|_2^2 \\ &\text{s.t.} \quad \left\| \boldsymbol{\alpha}_{ijk} \right\|_0 \le K, \quad \forall i,j,k \end{aligned} \tag{4}$$

We apply the K-SVD algorithm to estimate $\mathbf{D}_t$. After the dictionary is learned, the orthogonal matching pursuit (OMP) algorithm is used for sparse coding. In the other step, $\mathbf{D}_t$ and $\mathbf{A}_t$ are both fixed. The problem is simplified to:

$$\begin{aligned} \min_{\mathbf{X}} \ &\mathrm{TV}\left(\mathbf{X}_{t,M}\right) + \lambda_1 \left\| \mathbf{F}_t^u \mathbf{X}_{t,M} - \mathbf{Y}_{t,M} \right\|_2^2 + \\ &\lambda_2 \sum_{i,j,k} \left\| \mathbf{R}_{ijk}\mathbf{X}_{t,M} - \mathbf{D}_t \boldsymbol{\alpha}_{ijk} \right\|_2^2 \end{aligned} \tag{5}$$

The above optimization problem is solved using the split Bregman method [17].

Although all the frames in $\mathbf{X}_{t,M}$ can be reconstructed once the problem is solved at a certain t, only the result $\mathbf{X}_t$ is preserved for the final result. Other frames are used for the initialization of the problem at t+1.

Since the formulation (3) is non-convex, an appropriate initialization is crucial to obtain a good result. The sliding window (SW) technique is used to form the initial guess of **X** [2]. Due to the random undersampling pattern, it is hard to get a set of data covering the whole k-space within a few frames. Nevertheless, we still average some neighboring k-space data and make the zero filling reconstruction as the initial value. Compared with the direct zero filling method, sliding window can reduce the aliasing and thus provide a better estimation. The length of the sliding window is determined by the reduction factor and the motion speed. The K-SVD algorithm also needs an initialization for $\mathbf{D}_t$. When t = 1, the 3D overcomplete DCT dictionary is used for initialization. As we mentioned in the previous section, $\mathbf{D}_t$ is used for the initial value when training $\mathbf{D}_{t+1}$ in the following computations.

3. SIMULATION RESULTS

We use a set of dynamic cardiac cine data to validate the proposed method. The original data size is 256×256×25 (#PE×#FE×#frame). The cardiac region with a size of 128×128×25 is extracted for simulation. The Cartesian random down sampling pattern along the PE direction is employed. The central part of the k-space is fully sampled with a total of eight PE lines. The net reduction factor in this simulation is 4. Reconstruction using k-t FOCUSS is also performed for comparison [6].

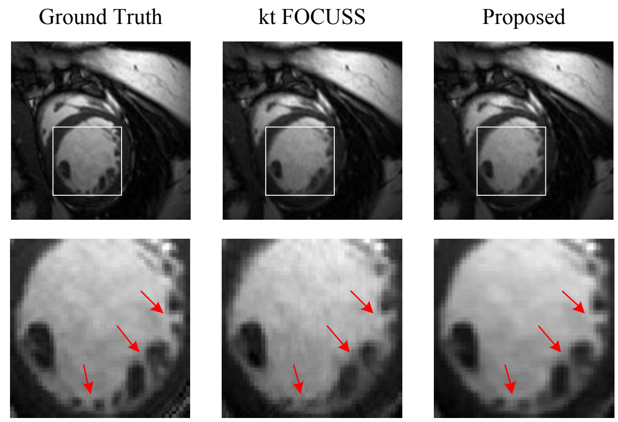

Fig. 1 Reconstructions of the fifth frame (Top row: images of reference, k-t FOCUSS and proposed. Bottom row: zoomed-in of select squares in reference, k-t FOCUSS and proposed)

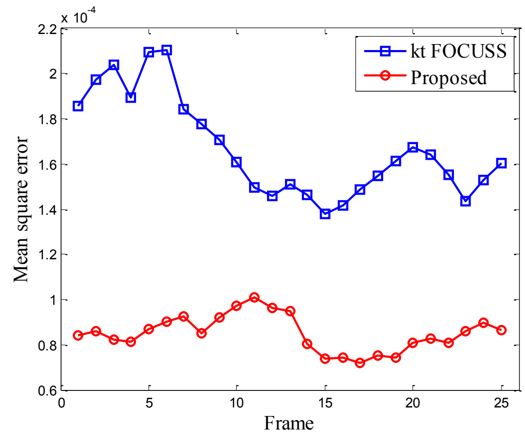

Fig. 2 Comparison of MSE in different frames

In the DL stage, the size of an atom (P) is $n_f \times n_p \times n_t$ = 4×4×5 = 80. The number of atoms (Q) is set to 400. The number of frames in one sub sequence (M) is 7. In K-SVD algorithm, the sparsity constraint is set to 5. As for the initial guess of **X**, the size of the sliding window is 3.

Fig. 1 shows the reconstruction results of the fifth frame. The zoomed-in figures of the cardiac region indicate that the proposed method preserves more details in the image. In order to provide a comparison of the reconstruction errors, Fig. 2 depicts the mean-squared error (MSE) between the reconstruction and the reference frame by frame. It shows that the proposed method achieves a lower MSE for all frames.

Fig. 3 presents the temporal profile at a fixed position in the frequency encoding direction. Although k-t FOCUSS captures main motion features, it loses some temporal variations. In contrast, the proposed method preserves more details than k-t FOCUSS. To further demonstrate the ability of the proposed method on capturing abrupt motions, Fig. 4 depicts the intensity curve of a single voxel located at (48, 45) over time. The result of proposed method is closer to the true one than that of k-t FOCUSS especially when the temporal variation is rapid.

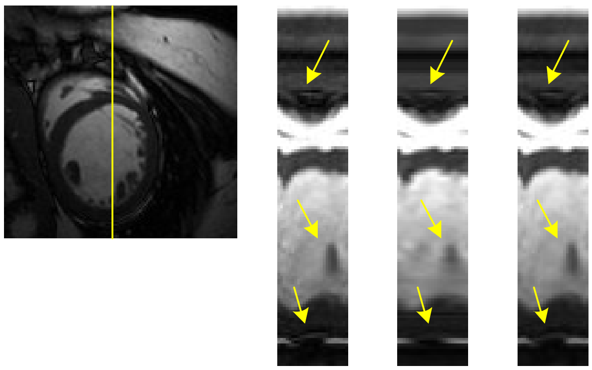

Fig. 3 Comparison of temporal profiles

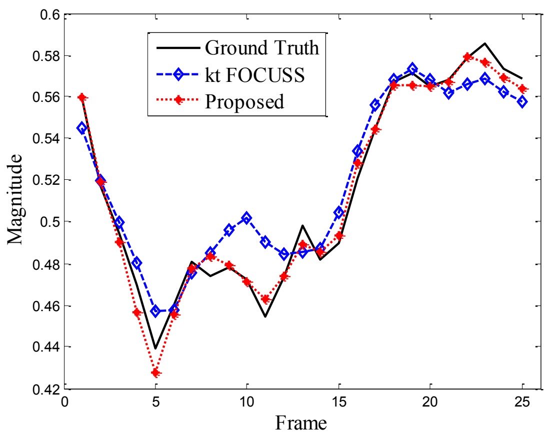

Fig. 4 Comparison of temporal intensity changes of a single voxel

4. CONCLUSION

In this paper, we propose a patch-based DL model for dMRI reconstruction. The main concept is to adopt a set of temporal-dependent dictionaries with three-dimensional atoms that are adaptively trained to provide sparse representations for compressed sensing reconstruction. The sparse representation is obtained by dividing the image sequence into three-dimensional overlapping patches and sparsifying these patches using the learned dictionaries. This patch-based DL model is able to capture local structure spatially and temporally. The simulation shows encouraging results that the proposed method can preserve both spatial structures and temporal variations.

5. ACKNOWLEDGMENT

This work is supported in part by the National Science Foundation CBET-0846514. The authors would like to thank Dr. Jong Ye for making the cardiac cine data available online (http://bisp.kaist.ac.kr/ktFOCUSS.htm).

6. REFERENCES

[1] J. Tsao and S. Kozerke, "MRI Temporal Acceleration Techniques," *J. Magn. Reson. Imag.*, vol. 36, no. 3, pp. 543-560, 2012.

[2] J. A. d'Arcy, D. J. Collins, I. J. Rowland, et al., "Applications of Sliding Window Reconstruction with Cartesian Sampling for Dynamic Contrast Enhanced MRI," *NMR Biomed.*, vol. 15, no. 2, pp. 174-183, 2002.

[3] 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, 2003.

[4] M. Lustig, J. M. Santos, D. L. Donoho, et al., "k-t SPARSE: High Frame Rate Dynamic MRI Exploiting Spatiotemporal Sparsity," *Proc. 14th ISMRM*, pp. 2420, 2006.

[5] U. Gamper, P. Boesiger and S. Kozerke, "Compressed Sensing in Dynamic MRI," *Magn. Reson. Med.*, vol. 59, no. 2, pp. 365-373, 2008.

[6] H. Jung, K. Sung, K. S. Nayak, et al., "k-t FOCUSS: A General Compressed Sensing Framework for High Resolution Dynamic MRI," *Magn. Reson. Med.*, vol. 61, no. 1, pp. 103-116, 2009.

[7] D. Liang, E.V.R. Dibella, R.R. Chen, et al., "k-t ISD: Dynamic Cardiac MR Imaging Using Compressed Sensing with Iterative Support Detection," *Magn. Reson. Med.*, vol. 68, no. 1, pp. 41-53, 2012.

[8] 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, 2006.

[9] M. Protter and M. Elad, "Image Sequence Denoising via Sparse and Redundant Representations," *IEEE Trans. Image Process.*, vol. 18, no. 1, pp. 27-36, 2009.

[10] I. Tošić and P. Frossard, "Dictionary Learning," *IEEE Signal Process. Mag.*, vol. 28, no. 2, pp. 27-38, 2011.

[11] X. Ye, Y. Chen, and F. Huang, "A Novel Method and Fast Algorithm for MR Image Reconstruction with Significantly Under-sampled Data," *Inverse Problems Imag.*, vol. 4, no. 2, pp. 223-240, 2010.

[12] 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, 2011.

[13] M. Doneva, P. Börnert, H. Eggers, et al., "Compressed Sensing Reconstruction for Magnetic Resonance Parameter Mapping," *Magn. Reson. Med.*, vol. 64, no. 4, pp. 1114-1120, 2010.

[14] S. G. Lingala and M. Jacob, "A Blind Compressive Sensing Framework for Accelerated Dynamic MRI," *Proc. 9th IEEE Int. Symp. Biomed. Imag. (ISBI)*, Barcelona, pp. 1060-1063, 2012.

[15] S. P. Awate and E.V.R. DiBella, "Spatiotemporal Dictionary Learning for Undersampled Dynamic MRI Reconstruction via Joint Frame-based and Dictionary-based Sparsity," *Proc. 9th IEEE Int. Symp. Biomed. Imag. (ISBI)*, Barcelona, pp. 318-321, 2012.

[16] 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, 2006.

[17] T. Goldstein and S. Osher, "The Split Bregman Method for L1-Regularized Problems," *SIAM J. Imag. Sci.*, vol. 2, no.2, pp. 323-343, 2009.