DYNAMIC MAGNETIC RESONANCE IMAGING USING COMPRESSED SENSING WITH SELF-LEARNED NONLINEAR DICTIONARY (NL-D)

Ukash Nakarmi[1], *Yanhua Wang*[1, 2], *Jingyuan Lyu* [1] *and Leslie Ying*[1, 3]
{ukashnak, yanhuaw, jlv27, leiying}@buffalo.edu
[1] Department of Electrical Engineering, [3] Department of Biomedical Engineering,
State University of New York at Buffalo, Buffalo, NY
[2] School of Information and Electronics, Beijing Institute of Technology, Beijing, P. R. China

ABSTRACT

Compressed Sensing (CS) is a new paradigm in signal processing and reconstruction from sub-nyquist sampled data. CS has shown promising results in accelerating dynamic Magnetic Resonance Imaging (dMRI). CS based approaches hugely rely on sparsifying transforms to reconstruct the dynamic MR images from its undersampled k-space data. Recent developments in dictionary learning and nonlinear kernel based methods have shown to be capable of representing dynamic images more sparsely than conventional linear transforms. In this paper, we propose a novel method (**NL-D**) to represent the dMRI more sparsely using self-learned nonlinear dictionaries based on kernel methods. Based on the proposed model, a new iterative approach for image reconstruction relying on pre-image reconstruction is developed within CS framework. Simulation results have shown that the proposed method outperforms the conventional CS approaches based on linear sparsifying transforms.

Index Terms **—** Compressed Sensing, Dictionary Learning, Non Linear Methods, Kernel Methods

1. INTRODUCTION

Dynamic Magnetic Resonance Imaging (dMRI) has series of images that carry the spatial and temporal information of the dynamic subject under consideration. These images not only characterize the spatial structures, but also capture the kinetic information [1]-[5] of a dynamic subject. However, the long data acquisition time in dMRI limits the achievable spatiotemporal resolution and thus compromises image quality. A recent paradigm on sub-nyquist sampling and signal reconstruction, Compressed Sensing (CS) has proven to be a powerful tool in accelerating data acquisition process in MRI [6]-[8] by exploiting the *a priori* information about the data sparsity under some sparsifying transforms. These sparsifying transforms play a vital role in the feasibility of CS reconstruction of MR images from under sampled k space data. CS in dMRI exploits the high correlation between inter frames along the temporal direction to find appropriate sparsifying transforms. Typical sparsifying transforms prevailed in dMRI are linear transforms such as Fourier [7-9], wavelet [6], principal component analysis (PCA) [8]. Another sparsifying approach, the so-called dictionary learning methods have also been explored in various literatures [10] [11]. However, all these methods are based on linear transforms and hence might not be able to capture nonlinear inter frame correlation of dMRI.

Nonlinear features in dMRI motivate the use of nonlinear sparsifying transforms that can effectively capture the nonlinear temporal correlation and thus allow sparser representations. Kernel Principal Component Analysis (KPCA) is a tractable nonlinear generalization of PCA [12] – [14] designed to capture such nonlinearities in the data. The key principle in KPCA method is to map the original data from the low dimensional input space to a higher dimensional space, so called feature space, through some nonlinear map and perform linear methods there. Recent works in various signal-processing applications such as data classification, feature recognition, active shape models and face recognition have shown promising results using KPCA to capture the underlying nonlinearities of the data [15] - [17]. While finding appropriate maps from input space to high dimensional feature space is of primary concern in such applications, implementation of KPCA methods in MRI context is even more challenging in a sense that, in MRI we are strictly concerned about the data in the input space rather than only on its characteristics in the feature space. The reverse mapping of data from feature space to input space, the so-called pre-image problem [14] [18] [19] is of equal importance in MRI and CS approaches. Kernel based CS reconstruction methods have been proposed in [20]-[22]. Recent works [23]-[25] investigate the application of kernel based CS in MRI. These approaches have shown superior results than conventional linear CS methods.

In this paper, we integrate the kernel based nonlinear dictionary approach within CS framework to sparsely represent the nonlinear features of temporal frames in dMRI. We propose a polynomial kernel based method to find a sparse representation of dynamic MR images from the under-sampled k-space data. Based on this model, a novel compressed sensing dMRI method with self-learned nonlinear dictionary (NL-D) is formulated. It is worth noting that although both [25] and our method use kernel PCA, [25] formulates the problem with a low-rank constraint in the feature space where the rank has to be specified, while our method enforces a sparsity constraint in the feature space with a learned dictionary. The rest of the paper is structured as follows. In section 2, we present our proposed method and elaborate various steps involved. Section 3 provides the simulation results for muscle arterial spin labeled (ASL) perfusion data and comparison with conventional linear CS-PCA method, and finally section 4 concludes the paper.

2. THEORY AND METHODS

The key idea in the proposed method is based on the argument that we can characterize the underlying nonlinear structures of dynamic MR images by low dimensional embedding in the higher dimensional feature space. The fundamental induction from this argument is that, dynamic images can be represented more sparsely using nonlinear dictionary such that the sparsity constrained CS reconstruction is more accurate. Given the undersampled k-space data $\mathbf{y}$, the reconstruction problem can be formulated as:

$$\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{F}_{\mathrm{u}}\mathbf{x}\|_2^2 + \eta_1 \|\boldsymbol{\tau}\|_1 , \qquad (1)$$

where $\mathbf{F}_u$ is undersampling Fourier operator, $\mathbf{x}$ is the desired dynamic image, $\boldsymbol{\tau}$ is the sparse coefficients with nonlinear dictionary and η_1 is regularization parameter. The nonlinear dictionary is formed using training data and projection into the feature space. The optimization problem in Eq. (1) is solved using a pre-image formulation and the iterative soft thresholding method. The process involved in the proposed method can be described in following 3 distinct steps: (I) non-linear dictionary learning, (II) sparsity enforcement, and (III) data consistency enforcement. The schematic of proposed method is illustrated in the Fig. 1.

2.1. Non-Linear Dictionary Learning

The non-linear dictionary is learned from the training data obtained from low-resolution dynamic images using kernel principal component analysis (KPCA) [13]. Low-resolution dynamic images are obtained from a few central k-space lines. A set of T training signals $\mathbf{p}_t$, $t = 1, 2, \dots T$ are formed from the low-resolution dynamic images. Each of the training signal $\mathbf{p}_t$ corresponds to the temporal variation of a particular spatial location as shown in step 1 of the Fig. 1. To find the corresponding nonlinear dictionary from these training signals, they are projected from the original input space to the high dimensional feature space. A principal component (PC) in a feature space serves as a dictionary element and can be represented as, $\mathbf{V}=\sum_{t=1}^{T} \alpha_t \overline{\phi}\,(\mathbf{p}_t)$, where $\overline{\phi}(\mathbf{p}_t) = \phi(\mathbf{p}_t) - \sum_{t=1}^{T} \phi(\mathbf{p}_t)\,/T$ represents the mapping of centered training data in the feature space, $\phi : \chi \rightarrow H$ is the nonlinear map from the low dimensional input space χ to a high dimensional feature space H, and α_t is the representation coefficient. However, since the mapping function ϕ is not known explicitly, we use KPCA to compute the representation dictionary. A $T \times T$ kernel matrix $\mathbf{K}_p$ is formed using the training data as:

$$\mathbf{K}_p = \begin{bmatrix} k(\mathbf{p}_1, \mathbf{p}_1) & k(\mathbf{p}_1, \mathbf{p}_2) & \cdots & k(\mathbf{p}_1, \mathbf{p}_\mathrm{T}) \\ k(\mathbf{p}_2, \mathbf{p}_1) & k(\mathbf{p}_2, \mathbf{p}_2) & \dots & k(\mathbf{p}_2, \mathbf{p}_\mathrm{T}) \\ \vdots & & \ddots & \vdots \\ k(\mathbf{p}_\mathrm{T}, \mathbf{p}_1) & k(\mathbf{p}_\mathrm{T}, \mathbf{p}_2) & \cdots & k(\mathbf{p}_\mathrm{T}, \mathbf{p}_\mathrm{T}) \end{bmatrix}, \qquad (2)$$

where $k(.,.)$ is a kernel function. In particular, we use polynomial kernel function defined by $k(\mathbf{x}_i, \mathbf{x}_j) =$

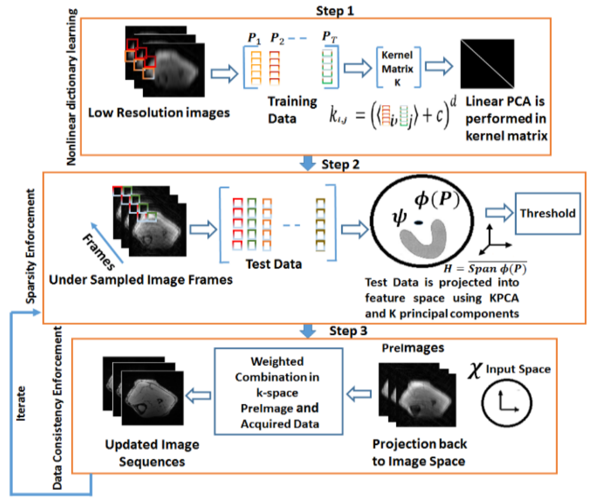

Fig. 1. Schematic of the proposed method

$(\langle \mathbf{x}_i, \mathbf{x}_j \rangle + c)^d$, where c is a constant and d is the order of polynomial. The centered kernel matrix of Eq. (2) is $\mathbf{K}_p^c = \mathbf{K}_p - \mathbf{1}_T \mathbf{K}_p - \mathbf{K}_p \mathbf{1}_T + \mathbf{1}_T \mathbf{K}_p \mathbf{1}_T$, where $\mathbf{1}_T$ is a $T \times T$ matrix with all its elements equal to $1/T$. The linear PCA is then performed in the feature space by solving the following eigenvalue problem

$$\mathbf{K}_P^c\ \boldsymbol{\alpha} = \lambda\, \boldsymbol{\alpha}\, , \qquad (3)$$

where $\sqrt{\lambda}$ is the length of PC and $\boldsymbol{\alpha} = [\alpha_1\ \alpha_2 \dots \alpha_T]^T$ are the representation coefficients.

2.2. Sparsity Enforcement

Similarly as in step 1 for training signals, the test signal vectors for each spatial location are formed using the images obtained from the undersampled k-space data as shown in step 2 of Fig. 1. Each test signal represents the temporal variation of the image at a particular spatial location. Letting such test signal denoted by $\mathbf{x}$, for each test signal, a kernel vector is calculated using

$$\mathbf{k}_{xp} = [k(\mathbf{p}_1, \mathbf{x})\ k(\mathbf{p}_2, \mathbf{x}) \dots k(\mathbf{p}_T, \mathbf{x})]^T\ , \qquad (4)$$

and the elements of the centered kernel vector is calculated using $\mathbf{k}_{xp}^c(t) = \mathbf{k}_{xp}(t) - \frac{1}{T}\sum_{i=1}^{T} \mathbf{K}_P(t,i) - \frac{1}{T}\sum_{i=1}^{T} \mathbf{k}_{xp}(i) + \frac{1}{T^2}\sum_{k=1}^{T}\sum_{i=1}^{T} \mathbf{K}_p(k,i)$. For a test signal $\mathbf{x}$, the projection of $\overline{\phi}\,(\mathbf{x})$ on to the k^{th} PC is computed using $\beta = \left(\mathbf{k}_{xp}^c\right)^T \boldsymbol{\alpha}^k$.

Similarly as in conventional PCA, we assume that $\overline{\phi}(\mathbf{x})$ can be sparsely represented using only K largest PCs in feature space as $\overline{\phi}(\mathbf{x}) \approx \sum_{k=1}^{K}\beta_k \mathbf{V}^k \approx \sum_{t=1}^{T}\overline{\gamma}_t \overline{\phi}(\mathbf{P}_t)$, where $\overline{\gamma_t} = \sum_{k=1}^{K}\beta_k\alpha_t^k$. The feature space map of the test signal $\mathbf{x}$, namely $\phi(\mathbf{x})$ is then computed as,

$$\phi(\mathbf{x}) = \overline{\phi}(\mathbf{x}) + \frac{1}{T}\sum_{t=1}^{T}\phi(\mathbf{p}_t) \approx \sum_{t=1}^{T}\gamma_t\phi(\mathbf{p}_t)\,, \qquad (5)$$

where $\gamma_t = \overline{\gamma_t} + \left(1 - \sum_{t=1}^{T}\overline{\gamma_t}\right)/T$. The sparse approximation of $\phi(\mathbf{x})$ from Eq. (5) depends on the numbers of PCs used. This process is carried out for each of the test signal. Hence, $\boldsymbol{\tau}$ in Eq. (1) is defined as the vectorized projection of all test data over K principal components in feature space. After obtaining the sparse representation in the feature space, $\phi(\mathbf{x})$ needs to be projected back onto the original space (known as the pre-image problem). For odd order polynomial kernels, the pre-image [14] [18] is obtained as:

$$\mathbf{z} = \sum_{i=1}^{N} f_k^{-1}\left(\sum_{t=1}^{T}\gamma_t\, k(\mathbf{p}_t,\boldsymbol{\xi}_i)\right)\boldsymbol{\xi}_i\,, \qquad (6)$$

where f_k^{-1} is the inverse polynomial kernel function, $k(\boldsymbol{x}_i,\boldsymbol{x}_j) = f(<\boldsymbol{x}_i,\boldsymbol{x}_j>)$ and $\{\boldsymbol{\xi}_1,\boldsymbol{\xi}_2,\dots\boldsymbol{\xi}_N\}$ are any orthonormal basis of the input space.

2.3. Data Consistency Enforcement

In order to make the reconstruction consistent with measured k-space data at the sampled k-space locations, the k-space data is updated using the weighted combination of measured k-space data and k-space data obtained from the pre-image as

$$\hat{\mathbf{y}} = \frac{\mathbf{y}+\eta_2\mathbf{w}}{1+\eta_2}\,, \qquad (7)$$

where $\mathbf{w}$ is the Fourier transform of the pre-image $\mathbf{z}$, and η_2 is the weight. The updated dynamic images are then obtained by inverse Fourier transform of the updated k-space data. The sparsity enforcement and data consistency enforcement steps are iterated until convergence.

3. SIMULATIONS AND RESULTS

We used muscle ASL perfusion data on a calf muscle to evaluate the proposed method. *Data acquisition parameters*: TR/TE = 2.8/1.2 ms, flip angle = 5^o, FOV= 160 × 112 mm^2 and matrix size = 112 × 100 × 20 (#FE × # PE × # frames). 1-D random under-sampling pattern along PE was used frame by frame. The net reduction factor was 3.

Simulation Parameters: Low-resolution images are obtained using 5 central k-space lines to create a training data set. Number of training signals T = 1000. Polynomial kernel with c=1 and $d = 3$ was used and number of PCs (K) = 80. Soft thresholding values, η_1 and η_2 were tuned appropriately.

Fig. 2 shows the comparison of the proposed method with conventional linear PCA based CS reconstruction method. We show the reconstruction of frames 3 and frame 10. All computations were carried out on a DELL workstation with

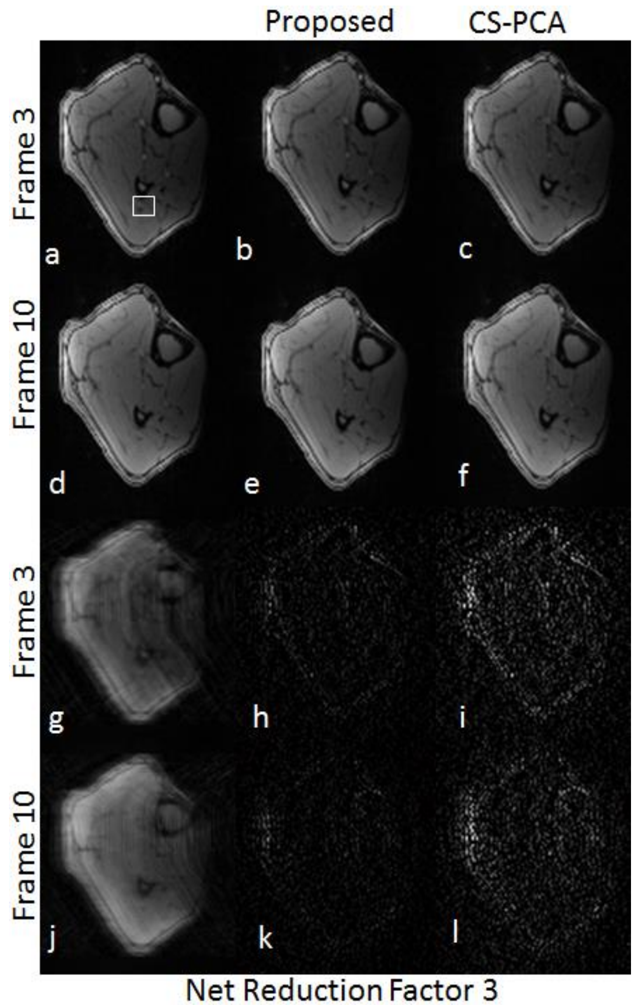

Fig.2. Comparison of different reconstructions. Reference: a & d, zero filling g & j, error maps h, i, k, l.

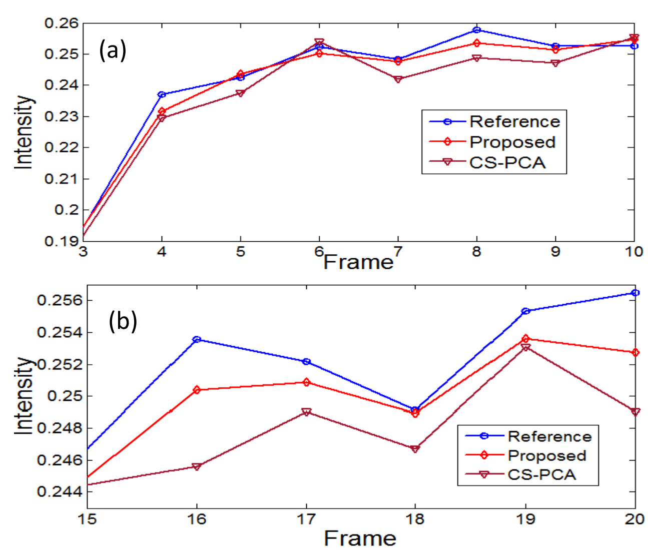

Fig. 3 Average Intensity curve of ROI. (a) Frames 3-10. (b) Frames: 15-20.

Intel(R) i7 3.40GHz CPU and 16GB RAM running MATLAB 2014. The reconstruction time is about 210s for the proposed method and 90s for CS-PCA. The reconstruction of lower frames are comparatively challenging than higher frames because of low signal to noise ratio (SNR). However, the higher frames are also of equal importance to capture the kinetic information. The proposed method outperforms the conventional linear CS-PCA method. The error maps strongly suggest that the proposed method is superior to the CS-PCA. Moreover, we can see the aliasing artifacts in the CS-PCA method.

In Fig. 3 (a) and (b), we present the average temporal curve of the selected region of interest (ROI) shown in Fig.2 (a). For better visualization, we depict the temporal curves of particular frames in two different figures, Fig. 3 (a) and (b). We can clearly observe that the temporal curves from the proposed method follows the reference temporal curve more closely and precisely. These results demonstrate that the proposed method is superior to conventional CS-PCA not only in image quality but also in preserving the rapid kinetic information.

4. CONCLUSION

In this paper, we developed a novel compressed sensing reconstruction method based on the kernel method and self-learned nonlinear dictionary. The proposed method integrates the principles of nonlinear kernel method into compressed sensing framework. We showed that the nonlinear features of temporal frames can be more efficiently and sparsely represented with self-learned nonlinear dictionary using kernel PCA. Simulation results have verified that the proposed method outperforms the conventional linear transform based CS methods. Proposed method preserves not only the spatial information but also the temporal kinetic information of dMRI images. It would be interesting to investigate the performance of updating the nonlinear dictionary iteratively, and to explore other optimization algorithms for improved convergence in future studies.

ACKNOWLEDGEMENT

This work is supported in part by the National Science foundation CBET-1265612.The authors would also like to thank Dr. Jie Zheng from Washington University in St. Louis for kindly providing the muscle ASL perfusion data.

REFERENCES

[1] Z.-P. Liang; P. C. Lauterbur, "An efficient method for dynamic magnetic resonance imaging," *IEEE Transactions on Medical Imaging*, vol.13, no.4, pp.677,686, Dec 1994.

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

[3] Z.-P. Liang, "Spatiotemporal imaging with partially separable functions," *Proc. 2007 IEEE ISBI*, pp. 988-991.

[4] J. Tsao and S. Kozerke, "MRI temporal acceleration techniques," J. Magn. Reson. Imag., vol. 36, pp. 543-560, 2012.

[5] 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, pp. 1031-1042, 2003.

[6] M. Lustig, J. M. Santos, D. L. Donoho, et al., "k-t SPARSE: High frame rate dynamic MRI exploiting spatiotemporal sparsity," *in Proc. ISMRM*, pp. 2420 , 2006.

[7] U. Gamper, P. Boesiger and S. Kozerke, "Compressed sensing in dynamic MRI," *Magn. Reson. Med.*, vol. 59, pp. 365-373, 2008.

[8] 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, pp. 103-116, 2009.

[9] 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, pp. 41-53, 2012.

[10] S. Ravishankar and Y. Bresler, "MR Image Reconstruction From Highly Undersampled k-Space Data by Dictionary Learning," *IEEE Transactions on Medical Imaging,* vol.30, no.5, pp.1028,1041, May 2011.

[11] Y. Wang and L. Ying. "Compressed sensing dynamic cardiac cine MRI using learned spatiotemporal dictionary." *IEEE Trans. Biomed. Eng.*, vol. 61, pp. 1109-1120, 2014.

[12] B. Schlkopf and A. J. Smola, "Learning with kernels: support vector machines, regularization, optimization, and beyond." MIT Press, Boston, 2001.

[13] B. Scholkopf, A. Smola, and K.R. Muller, "Kernel principal component analysis," *Proc. 1997 ICANN LNCS*, pp. 583-588.

[14] S. Mika, B. Schölkopf, A. Smola, K.L. Müller, M. Scholz, and G. Rätsch, "Kernel PCA and de-noising in feature spaces," *Adv. Neural. Inf. Process. Syst.*, vol. 11, pp. 536-542, 1999.

[15] S. Gao, I. W. Tsang, and L.-T. Chia, "Kernel sparse representation for image classification and face recognition," *in Proc. 2010 Euro. Conf. Comput. Vision*, pp. 1-14.

[16] Y. Chen, N. M. Nasrabadi, and T. D.Tran, "Hyperspectral image classification via kernel sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, pp. 217-231, 2013.

[17] Y. Zhou, K. Liu, R. E. Carrillo, K. E. Barner, and F. Kiamilev, "Kernel-based sparse representation for gesture recognition." *Pattern Recognition*, vol. 46, pp. 3208–3222, 2013.

[18] J.T.-Y Kwok, I.W. Tsang, "The pre-image problem in kernel methods," *IEEE Trans. on Neural Networks*, vol.15, no.6, pp.1517,1525, Nov. 2004.

[19] P. M. Rasmussen, T.J. Abrahamsen, K.H. Madsen, L.K. Hansen, "Nonlinear denoising and analysis of neuroimages with kernel principal component analysis and pre-image estimation", *NeuroImage*, vol. 60, pp. 1807-1818, April 2012.

[20] S. Gao, I. W. Tsang, and C. Liang-Tien, "Sparse representation with kernels," *IEEE Trans. Image Process.*, vol. 22, pp. 423-434, 2013.

[21] H. Qi and S. Hughes, "Using the kernel trick in compressive sensing: Accurate signal recovery from fewer measurements." *Proc. 2011 IEEE ICASSP*, pp. 3940-3943.

[22] F.P. Anaraki and S. Hughes, "Kernel compressive sensing." *in Proc. 2011 IEEE Int. Conf. Image Process.*, pp. 494-498.

[23] Y. Zhou, Y. Wang and L. Ying, "A kernel-based compressed sensing approach to dynamic MRI from highly undersampled data." *Proc.2013 IEEE ISBI*, pp. 310-313.

[24] J. F. M. Schmidt and S. Kozerke, "MR Image Reconstruction Exploiting Nonlinear Transforms,''*Proc. ISMRM*, pp. 746, 2014.

[25] Y. Wang and L. Ying, "Undersampled dynamic magnetic resonance imaging using kernel principal component analysis," *Proc. 2014 IEEE EMBC*, pp.1533-1536.