SPATIOTEMPORAL DICTIONARY LEARNING FOR UNDERSAMPLED DYNAMIC MRI RECONSTRUCTION VIA JOINT FRAME-BASED AND DICTIONARY-BASED SPARSITY

Suyash P. Awate

Scientific Computing and Imaging Institute, University of Utah.

Edward V. R. DiBella

Utah Center for Advanced Imaging Research, University of Utah.

ABSTRACT

Image reconstruction using compressed sensing relies on sparse representations of signals in some dictionary. Current state-of-the-art dictionary-learning methods are designed for spatial images and fail to systematically generalize to dynamic imaging scenarios where the spatiotemporal data, and thereby the spatiotemporal dictionary atoms, exhibit joint coherence in space and time leading to low rank. This paper proposes a novel method for learning *low-rank spatiotemporal dictionaries*. While leading compressed-sensing reconstruction methods employ either l_1 analysis or synthesis approaches using mathematical frames (e.g. overcomplete wavelets), approaches using dictionary learning (very recent) ignore the frame-based l_1-sparsity constraints. This paper proposes a novel method *combining frame-based l_1 analysis with spatiotemporal-dictionary* based sparsity (related to l_1 synthesis). The results demonstrate improved reconstructions, on simulated and clinical highly-undersampled dynamic images, using the combined approach.

Index Terms— Reconstruction, dynamic MRI, undersampling, compressed sensing, dictionary learning.

1. INTRODUCTION AND RELATED WORK

State-of-the-art methods for signal reconstruction from undersampled magnetic resonance imaging (MRI) rely on principles in compressed sensing, which allow complete signal recovery despite sub-Nyquist sampling rates if (i) the signal has a sparse linear representation (in some mathematical frame or a *dictionary*) and (ii) there is low coherence between the sensing basis (i.e. the Fourier basis in MRI) and the representation frame [1]. While the early dictionaries were mathematical frames, current research focusses on *learning*/adapting the dictionary for the (class of) signals being reconstructed [2].

Leading dictionary-learning methods [2], including K-SVD, learn a set of atoms representing spatial patterns for spatial images. However, dynamic imaging produces *spatiotemporal images* that exhibit *joint coherence in space and time*. Indeed, [3] exploits this space-time coherence for undersampled-MRI reconstruction via a low-rank constraint on the reconstructed image. However, state-of-the-art dictionary-learning methods [2] fail to exploit this coherence. This paper proposes a novel formulation, and algorithm, for learning *low-rank spatiotemporal dictionaries*.

While leading compressed-sensing reconstruction methods [1, 2] employ either l_1 analysis or synthesis approaches using frames (e.g. overcomplete wavelets), approaches using dictionary learning (very recent [4]) ignore frame-based l_1 sparsity. This paper proposes a novel approach *combining frame-based l_1 analysis with spatiotemporal-dictionary* based (related to l_1 synthesis) sparse representation. The results demonstrate improved reconstructions, on simulated and clinical data, using the combined approach. Moreover, while [4] includes data fidelity in the Lagrangian leading to a least-squares reconstruction (given the dictionary), the proposed approach includes fidelity as a constraint based on the noise variance (typically easier to tune than the Lagrange weight parameter) leading to constrained l_1-analysis optimization.

2. SPATIOTEMPORAL DICTIONARY LEARNING

This section proposes a novel formulation for learning spatiotemporal dictionaries enforcing low-rank constraints and proposes an efficient algorithm for optimization.

Consider a *spatiotemporal patch* of size $n_S \times n_T$, where n_S is the number of pixels chosen in the space dimension (a spatial size of $2\times2 \Rightarrow n_S = 2^2$) and n_T is the number of timepoints in the data. Let X be a matrix comprising spatiotemporal data where the i-th column $x_i \in \mathbb{R}^{n_S n_T}$ is a vectorized spatiotemporal patch. Let D be a matrix representing the dictionary where the i-th column $d_i \in \mathbb{R}^{n_S n_T}$ is a vectorized spatiotemporal atom. Let C be a sparse matrix of coefficients where the i-th column c_i is such that $Dc_i \approx x_i$. Let $\mathbb{M}^{\text{LowRank}}$ be the space of low-rank $n_S \times n_T$ matrices, with rank less than $\min(n_S, n_T)$. We formulate spatiotemporal dictionary learning as the constrained optimization problem:

$$D^{\text{opt}} \triangleq \arg\min_D \min_C \|X - DC\|_2^2 \text{ s.t. } \forall i, \|c_i\|_0 \leq T_0, \\ \text{and } \forall i, \mathcal{U}(d_i, n_S, n_T) \in \mathbb{M}^{\text{LowRank}}, \quad (1)$$

The authors gratefully acknowledge funding via NIH R01 EB00177.

where operator $\mathcal{U}(\cdot, n_S, n_T)$ undoes/reverses vectorization reshaping a vectorized spatiotemporal atom d_i into a $n_S \times n_T$ matrix. Our optimization scheme is: (i) initialize D, (ii) fix D and optimize C, (iii) fix C and optimize D, (iv) repeat the last two steps until convergence. Details appear next.

Given data X, we *initialize the atoms* d_i in D as a chosen subset of the spatiotemporal patches x_i. Dictionary-based reconstruction works best when the atoms d_i have low mutual coherence [5] or when the columns of the Gram matrix D^*D are reasonably sparse [6] leading to rapidly-decaying coefficients for sparsely-represented signals. Thus, dictionary learning methods need to ensure these properties. We initialize the dictionary D using farthest-point clustering [7] to cluster the spatiotemporal patches x_i into a maximum number of clusters such that the coherence between any pair of cluster centers (i.e. atoms d_i^{init}) is less than μ (a free parameter).

Given D, we *update coefficients* C via sparse coding [2]:

$$C^{\text{opt}} \triangleq \arg\min_{C} \|X - DC\|_2^2 \text{ s.t. } \forall i, \|c_i\|_0 \leq T_0, \quad (2)$$

by using iteratively-reweighted least squares [5] independently for each spatiotemporal patch x_i.

Given C, we *update each dictionary atom* d_i successively as follows. Let c^i (superscript) denote the i-th *row* in C. Then, $\|X - DC\|_2^2 = \|E(i) - d_i c^i\|_2^2$, where the residual error matrix $E(i) \triangleq X - \sum_{j \neq i} d_j c^j$ is independent of d_i. In row c^i, the indices $\{k\}$ that have practically zero values, induced by sparsity constraints on C, correspond to data x_k for which atom d_i does *not* participate in representation. Thus, modifications to d_i become irrelevant for columns $\{k\}$ of the residual error matrix $E(i)$. Let $\Omega(i)$ be a projection operator eliminating columns with indices $\{k\}$. Then,

$$\begin{aligned}
&\|X - DC\|_2^2 = \|E(i) - d_i c^i\|_2^2 = \|E(i)\Omega(i) - d_i c^i \Omega(i)\|_2^2 \\
&= \|\widetilde{E} - d_i \tilde{c}^i\|_2^2 = \sum_m \|\tilde{e}_m - \tilde{c}_m^i d_i\|_2^2 \\
&= \sum_m \|\mathcal{U}(\tilde{e}_m, n_S, n_T) - \tilde{c}_m^i \mathcal{U}(d_i, n_S, n_T)\|_2^2, \quad (3)
\end{aligned}$$

where $\widetilde{E} \triangleq E(i)\Omega(i)$, $\tilde{c}^i \triangleq c^i \Omega(i)$, $\tilde{e}_m$ is the m-th column of $\widetilde{E}$, and $\tilde{c}_m^i$ is a scalar at the m-th column of row c^i. Thus,

$$\begin{aligned}
d_i^{\text{opt}} \triangleq \arg\min_{d_i} &\sum_m \|\mathcal{U}(\tilde{e}_m, n_S, n_T) - \tilde{c}_m^i \mathcal{U}(d_i, n_S, n_T)\|_2^2 \\
&\text{s.t. } \mathcal{U}(d_i, n_S, n_T) \in \mathbb{M}^{\text{LowRank}}. \quad (4)
\end{aligned}$$

This minimization problem involves a convex (quadratic) objective function and a convex constraint set $\mathbb{M}^{\text{LowRank}}$. We find the global minimum via projected gradient descent [8] on $\mathcal{U}(d_i, n_S, n_T)$ where: (i) the initialization is the *unconstrained* minimum $\sum_m \mathcal{U}(\tilde{e}_m, n_S, n_T) / \sum_m \tilde{c}_m^i$ and (ii) the projection on $\mathbb{M}^{\text{LowRank}}$ is given by the singular-value decomposition of $\mathcal{U}(d_i, n_S, n_T)$. We find that typically 2 iterations of gradient descent lead to convergence. We specify the low-rank constraint as the minimum fraction (we use 85%) of the energy in the singular-value spectrum that must be preserved. This constraint *regularizes the spatiotemporal patterns in atoms, making dictionary learning robust* to noise/artifacts.

3. RECONSTRUCTION VIA JOINT FRAME-BASED AND DICTIONARY-BASED SPARSITY

This section proposes a novel constrained-optimization formulation for reconstructing dynamic MRI by combining spatiotemporal-dictionary based sparsity (related to l_1 synthesis) with the frame-based l_1-analysis approach. Given undersampled data z, we define the reconstructed image u as:

$$\begin{aligned}
u^{\text{opt}} \triangleq \arg\min_{u} \min_{D,C} &\lambda \|\Psi u\|_1 + (1-\lambda) \sum_j \|R_j u - D c_j\|_2^2 \\
&\text{s.t. } \|\widehat{F} u - z\|_2^2 \leq \eta; \quad \forall j, \|c_j\|_0 \leq T_0; \\
&\text{and } \forall i, \mathcal{U}(d_i, n_S, n_T) \in \mathbb{M}^{\text{LowRank}}, \quad (5)
\end{aligned}$$

where $\lambda \in [0,1]$ is a weight parameter, Ψ is a tight-frame analysis transform (we use an overcomplete wavelet transform), $\widehat{F}$ is the undersampled Fourier transform, and η is the noise variance. The operator R_j extracts a spatiotemporal patch from u and vectorizes it. Note: because this paper reconstructs cardiac perfusion MRI (Section 5), motivated by underlying physiology [9], we consider D as real; so R_j also takes the magnitude at each index of the extracted vector.

We propose the following optimization scheme: (i) initialize: fix λ=1 and optimize u, (ii) set $\lambda \in [0,1]$ (iii) fix u and optimize D, C, (iv) fix D, C and optimize u, (v) repeat last two steps until convergence. Step (iii) leads to a spatiotemporal dictionary-learning problem that we solve using the proposed method in Section 2. Steps (i) and (iv) entail minimizing a non-smooth convex function (i.e. $\lambda\|\Psi u\|_1 + (1-\lambda)\sum_j \|R_j u - D c_j\|_2^2$) over a convex constraint set (i.e. $\|\widehat{F} u - z\|_2^2 \leq \eta$), which we solve using the fast-converging optimization scheme in [10] coupled with the reweighted l_1-analysis approach [1]. Typical compressed-sensing methods formulate the optimization problem in Lagrangian form (i.e. the basis-pursuit denoising problem [5]) that includes $\|\widehat{F} u - z\|_2^2$ in the objective function at the cost of a free weight parameter that is usually harder to tune than η.

4. VALIDATION USING SIMULATED DATA

This section validates the proposed methods for reconstruction and dictionary learning on a simulated spatiotemporal phantom. The phantom is a sequence of 2D images each comprising 2 ellipses (Figure 1(a)) whose intensities change smoothly over time (Figure 1(b)). Figure 1(c) shows the pseudo-radial acquisition (undersampling factor 13.4).

The sliding-window reconstruction (Figure 1(d)-(e)) is very noisy. The reconstruction using only a 3D overcomplete wavelet transform (λ = 1; Figure 1(f)-(g)) is also of low

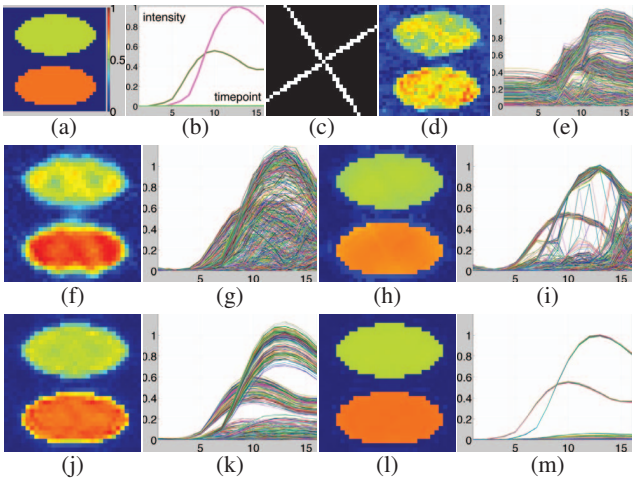

Fig. 1. **Spatiotemporal simulated-phantom reconstruction.** The phantom is a sequence of 16 2D images of size 32×32 pixels. **(a)** 10-th image from the sequence. **(b)** Time-varying intensities for the ellipses and the background (close to zero). **(c)** Pseudo-radial kspace acquisition pattern for 10-th timepoint; kspace center $\equiv$ image center; data acquired at bright pixels; other timepoints use a randomly-rotated pattern. **(d) Sliding-window** reconstruction. **(f)** Reconstruction using an **overcomplete 3D (space-time) wavelet** transform (Ψ); λ = 1. **(h)** Reconstruction using a **spatial dictionary** model (D) and the wavelet transform in (f); $\lambda = 0.5$. **(j)** Reconstruction using a **temporal dictionary** model (D) and the wavelet transform in (f); λ = 0.5. **(l)** Reconstruction using a **spatiotemporal dictionary** model (D) and the wavelet transform in (f); $\lambda = 0.5$. **(e),(g),(i),(k),(m)** show the temporal curves for all 32^2 pixels in space in (d),(f),(h),(j),(l), respectively.

quality. The next 3 results use a combination of the aforementioned wavelet transform with 3 different dictionary models (with λ = 0.5): spatial (Figures 1(h)-(i)), temporal (Figures 1(j)-(k)), and spatiotemporal (Figures 1(l)-(m)). These results demonstrate that coupling frame-based sparsity with dictionary-based sparsity can lead to significantly improved reconstructions. Spatial and temporal dictionary models improve reconstructions, but lead to errors around edges and in the temporal patterns. A spatiotemporal dictionary model produces an almost error-free reconstruction demonstrating the advantages of learning a spatiotemporal dictionary.

We now show the dictionaries learned during the reconstructions in Figure 1. All dictionaries contain a constant unit-norm atom absent from the figures. Figure 2(a) shows 6 spatial-dictionary atoms that capture patches around edges of the ellipses. Figures 2(b)-(c) show 2 temporal-dictionary atoms that capture temporal intensity patterns in the ellipses. Figures 3 shows a single low-rank spatiotemporal-dictionary atom (D contains 27 atoms) that captures both temporal and spatial intensity patterns as well as their inter correlation. Thus, it learns a more complex pattern underlying the data.

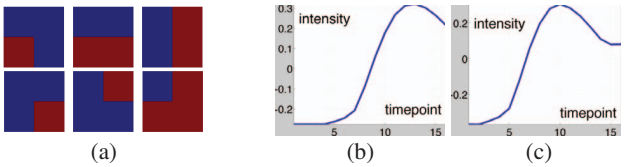

Fig. 2. **(a) 6 spatial-dictionary atoms** (size 2×2 pixels) capturing spatial edge patterns; red $\equiv$ 1, blue $\equiv$ 0 intensity value. **(b)-(c) 2 temporal-dictionary atoms** (size 1×16) capture the temporal intensity patterns in the 2 ellipses.

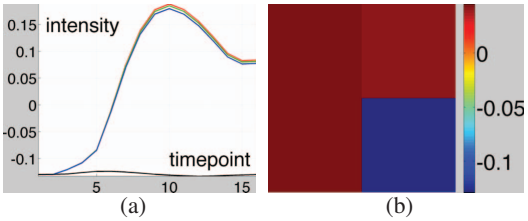

Fig. 3. **A single spatiotemporal-dictionary atom** (size $2 \times 2 \times 16$; $n_S = 4$, $n_T = 16$; rank 2). **(a)** 4 temporal patterns for 4 pixels in the space dimension. **(b)** Mean of the 4 temporal patterns, indicating the spatial pattern captured.

5. RESULTS USING CLINICAL DATA

This section shows results of the proposed methods for reconstruction and spatiotemporal dictionary learning on *clinical* contrast-enhanced cardiac perfusion MRI. The dataset is a fully-sampled, *but very noisy*, sequence of 29 2D images of size 192×96 (Figures 4(a)–(c)). We reconstruct images after simulated pseudo-radial acquisitions (Figure 4(d); undersampling factor 11.5) and use the fully-sampled data for qualitative assessment. Figure 4(e) shows the typical variation of root-mean-squared difference with λ suggesting that roughly-equal contributions ($\lambda \in [0.5, 0.7]$) from dictionary-based and wavelet-based sparsity produce desirable results.

The sliding-window reconstruction (Figure 4(f)-(g)) retains most of the noise in the images and the temporal intensity patterns. The reconstruction using a 3D overcomplete wavelet (λ = 1) (Figure 4(h)-(i)) blurs edges significantly more than the dictionary-based reconstructions using temporal (Figure 4(j)-(k)) and spatiotemporal (Figure 4(l)-(m)) dictionaries. While using the temporal dictionary leads to sharper edges (sometimes slightly aliased) in the reconstruction, using the spatiotemporal dictionary balances detail preservation with spatial and temporal denoising; indeed, it has the *least variation in temporal patterns* over pixels. This can benefit clinical diagnostic applications (e.g. concerning cardiac ischemia) requiring reliability of the temporal perfusion characteristics of the contrast agent in the heart regions.

Figure 5 shows temporal-dictionary atoms learned during the reconstruction in Figure 4(j)-(k). The atom in Figure 5(b) reflects the contrast uptake curve obtained from the right-ventricular blood pool [9]. The atoms in Figures 5(c)-(d) are delayed, washed-out, and dispersed version of the con-

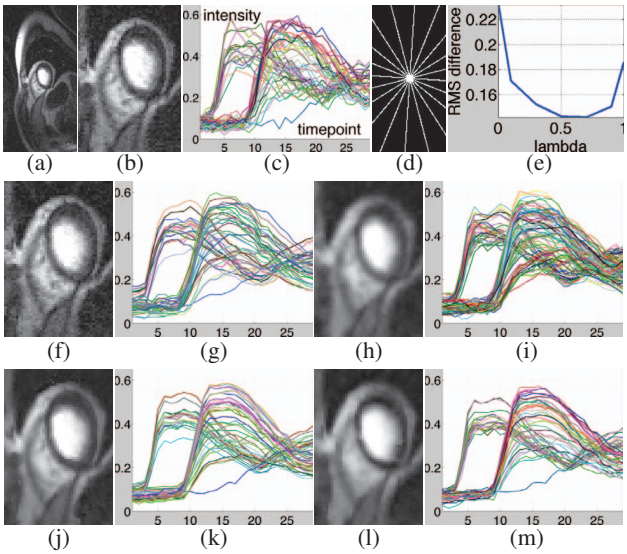

Fig. 4. *Please see the electronic document for clarity.* **Spatiotemporal cardiac-perfusion MRI reconstruction. (a)** Fully-sampled image at the 14-th timepoint. **(b)** Zoomed-in view of the heart region in (a). **(d)** Simulated pseudo-radial kspace sampling for the 14-th timepoint; other timepoints use a randomly-rotated sampling pattern; 10 lines; undersampling factor 11.5. **(e)** Root-mean-squared magnitude difference between the reconstructed and fully-sampled spatiotemporal images divided by norm of the fully-sampled image. **(f) Sliding-window** reconstruction. **(h)** Reconstruction using a **3D (space-time) overcomplete wavelet** transform (Ψ); $\lambda = 1$. **(j)** Reconstruction using a **temporal dictionary** model (D) and the wavelet transform (Ψ) in (h); $\lambda = 0.5$. **(l)** Reconstruction using a **spatiotemporal dictionary** model (D) and the wavelet transform (Ψ) in (h); $\lambda = 0.5$. **(c),(g),(i),(k),(m)** show temporal intensity patterns within both ventricles in (b),(f),(h),(j),(l), respectively.

trast uptake curve, as expected by the parametric physiological perfusion model [9] for temporal magnitude curves. The atom in Figure 5(d) models the global uptake of the contrast, reflecting a long delay. Thus, these atoms are consistent with the underlying physiological process.

Figure 6 shows low-rank spatiotemporal dictionary atoms (learned during the reconstruction in Figure 4(l)-(m)) that capture temporal and spatial patterns as well as their inter correlations. The spatiotemporal dictionary indeed captures complex patterns outside the parametric space.

Conclusions: This paper has proposed novel formulations and algorithms for (i) low-rank spatiotemporal dictionary learning (Section 2) and (ii) compressed-sensing reconstruction of dynamic MRI combining frame-based l_1-analysis with dictionary-based sparsity (related to l_1 synthesis) (Section 3). Results on simulated and clinical dynamic MRI demonstrate the advantages of the proposed methods.

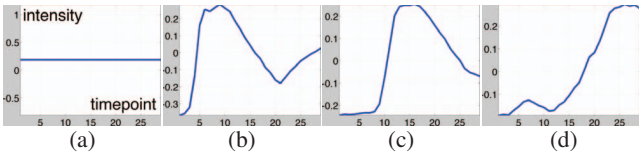

Fig. 5. **4 temporal-dictionary atoms** (size 1×29) for cardiac-perfusion MRI.

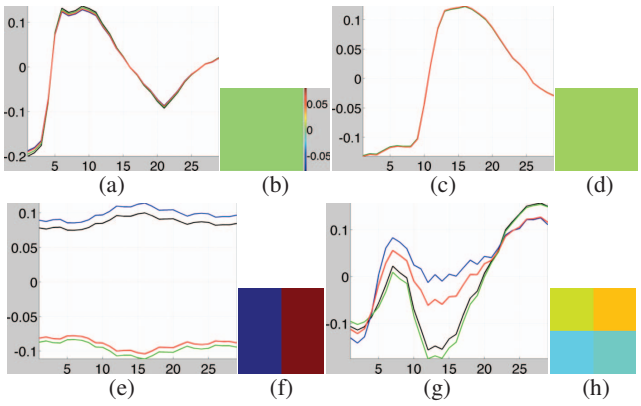

Fig. 6. **4 spatiotemporal-dictionary atoms** (size $2 \times 2 \times 29$; $n_S = 4$, $n_T = 29$) for cardiac-perfusion MRI. **(a),(c),(e),(g)** each show 4 temporal patterns for 4 pixels in the space dimension. **(b),(d),(f),(g)** each show the means of temporal patterns in (a),(c),(e),(g), respectively, indicating the spatial pattern captured. Ranks for the 4 atoms: 1,1,1, and 2.

6. REFERENCES

[1] E. Candes and M. Wakin, "An introduction to compressive sampling," *IEEE Sig. Proc. Mag.*, pp. 21–30, 2008.

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

[3] J. Haldar and Z. Liang, "Low rank approximations for dynamic imaging," in *IEEE Int. Symp. Bio. Imag.*, 2011, pp. 1052–1055.

[4] S. Ravishankar and Y. Bresler, "MR image reconstruction from highly undersampled k-space data by dictionary learning," *IEEE Trans. Med. Imag.*, vol. 30, pp. 1028–1041, 2011.

[5] D. Donoho, M. Elad, and V. Temlyakov, "Stable recovery of sparse overcomplete representations in the presence of noise," *IEEE. Trans. Info. Theory*, vol. 52, no. 1, pp. 6–18, 2006.

[6] E. Candes, Y. Eldar, D. Needell, and P. Randall, "Compressed sensing with coherent and redundant dictionaries," *Appl. and Comp. Harmonic Anal.*, vol. 31, no. 1, pp. 59–73, 2010.

[7] T. Gonzalez, "Clustering to minimize the maximum intercluster distance," *Theor. Comp. Sci.*, vol. 38, pp. 293–306, 1985.

[8] D. Bertsekas, *Nonlinear programming*, Athena Scient., 2008.

[9] S. P. Awate, E.V.R. DiBella, T. Tasdizen, and R. Whitaker, "Model-based image reconstruction for dynamic cardiac perfusion MRI from sparse data," in *Proc. IEEE Eng. Med. Biol. Soc.*, 2006, pp. 936–41.

[10] Y. Nesterov, "Smooth minimization of non-smooth functions," *Math. Program., Ser. A*, vol. 103, pp. 127–152, 2005.