Harnessing Group-Sparsity Regularization for Resolution Enhancement of Lung 4D-CT

Arnav Bhavsar, Guorong Wu, and Dinggang Shen

Department of Radiology and BRIC, University of North Carolina at Chapel Hill

Abstract. A critical concern with lung 4D-CT is the low superior-inferior resolution, due to the consideration of radiation dose. We propose a resolution enhancement approach that reconstructs missing intermediate slices by exploiting the idea that information lost in one respiratory phase can be found in others, according to the complimentary nature of inter-phase information. Our approach is based on a patch-based framework that explores the role of group-sparsity involving groups of similar neighbouring patches. We discuss the regularizing role of group-sparsity, which helps in reducing the effect of noise and enables better enhancement of anatomical structures. Our results positively demonstrate the potential of group-sparsity for 4D-CT resolution enhancement.

1 Introduction

4D-CT plays a crucial role in radiation therapy for lung cancer. Unlike 3D-CT under free breathing, 4D-CT provides a more accurate estimate of the lung motion across respiratory phases. This helps to better localize the moving structures in the lung.

However, to control radiation dose in 4D-CT, usually a reduced number of slices are acquired, which results in low superior-inferior resolution. This can adversely affect the image quality [1] due to false apparent vessel discontinuities, shape distortions etc., which makes the assessment of tumor and vessel structures difficult [7]. This compromises the potential of 4D-CT for providing accurate structural information. Resolution enhancement here aims at addressing this important limitation.

Resolution enhancement, often known as super-resolution (SR), involves retrieving lost high-frequency sampling information (e.g. fine structures) [2,3]. Typically, SR approaches exploit multiple observations with relative sub-pixel motion [2] or a large dictionary of local patches from an off-line high-resolution (HR) image dataset [3,4].

For 4D-CT, it is difficult to rely on accurate registration due to poor superior-inferior resolution. While there have been attempts for correspondence-based interpolation [5] or super-resolution via motion-estimation [6], in general, registration-based method can be quite inaccurate, as demonstrated in [7]. Also, availability of large off-line HR datasets for constructing dictionaries is infeasible. Nevertheless, one can exploit motion-induced information in a patch-based framework without the need for registration and large HR data. Due to lung motion, local image information across respiratory phases is complimentary, which one can capture in a dictionary of patches [7,8].

We propose a group-sparsity-based approach to integrate such local inter-phase local information. Our approach involves multiple dictionaries that consider spatial neighbourhood and similarity. Our work is closely related to the sparse representation method

K. Mori et al. (Eds.): MICCAI 2013, Part III, LNCS 8151, pp. 139–146, 2013.

in [7], which, however, does not employ any neighbourhood constraints. Accordingly, relatively large-scale patches have to be used in [7], in order to resist noise and other artifacts. However, a large scale results in over-smoothing. On the other hand, a smaller scale to improve localization could result in an error-prone noisy reconstruction.

Such a behaviour is an example of 'noise-structure trade-off' [2]. From a clinical view, structural accuracy and localization is important for assessment of shape/extent of tumors and anatomical structures, which is vital in radiation therapy. It is also crucial that noise-effects and artifacts are minimized during reconstruction. Thus, the clinical importance of mitigating the 'noise-structure trade-off' can be clearly appreciated.

In this regard, we harness the regularizing potential of group-sparsity over spatially neighbouring patches. This enables a constructive use of smaller patch-scale, which helps in enhancing local structures, while still being robust to noise effects and artifacts.

Thus, our work contributes in: 1) Advancing the recent and clinically important area of 4D-CT resolution enhancement in a different framework than traditional SR methods; 2) Exploring the role of group-sparsity for image reconstruction; 3) Exploiting neighbourhood constraints for better structural enhancement.

2 Role of Group-Sparsity in Regularization

In many estimation problems, multiple factors (tasks) can be related. In such cases, the sparse representation problem can be expressed for a group of tasks as follows:

$$\hat{A} = \arg\min_{\alpha_l} \sum_{l=1}^{P} ||\mathbf{y}_l - D_l\alpha_l||_2^2 + \lambda||A||_{2,1}, \quad \text{where} \quad ||A||_{2,1} = \sum_{q=1}^{m} ||\mathbf{a}^q||_2 \quad (1)$$

Here, $\mathbf{y}_l$ is the observation and D_l is the dictionary for the l^{th} task. $A = [\alpha_1, .., \alpha_l, .., \alpha_P]$ is an $m \times P$ matrix whose columns are coefficient vectors related to the P different tasks. Each vector α_l describes the linear combination of atoms in D_l that best matches $\mathbf{y}_l$. $||A||_{2,1}$ is the $l_{2,1}$-norm, viz. an l_1-norm over the l_2-norms of the rows of A, where $\mathbf{a}^q$ denotes the q^{th} row of A. The $l_{2,1}$-norm enforces coefficient vectors for all tasks to have a similar sparsity structure, i.e., same locations of the zero elements. The intuition is that the related tasks employ the same (or similar) dictionary atoms.

Such an insight helps us to exploit the regularization potential of group-sparsity. We group patches into dictionaries, considering spatial neighbourhood and appearance similarity. The selected patches satisfy closeness with their neighbouring patches, resulting in smoothness, which, however, tends not to adversely affect the salient structures. Thus, group-sparsity plays a regularizing role towards mitigating the noise-structure trade-off. To our knowledge, only [10,11] employ group-sparsity for image reconstruction. While our method is very different, these works support such an intuition.

3 Methodology

Given 4D-CT data $I = \{I_i(s)|i = 1, ..., L; s = 1, ..., S\}$ (with L phases and S slices), we estimate a slice between $I_i(s)$ and $I_i(s+1)$ for the i^{th} phase, in a patch-wise manner [7]. We consider patches from $I_i(s)$ and $I_i(s+1)$ to form the observation vectors.

Ours is a two-stage strategy, where the second stage uses the reconstructed slice from the first. We elaborate our method for the first step, and briefly indicate the modifications in the second. Patches are used as lexicographic vectors and, for simplicity, we do not use the subscript denoting the phase in which the slice is reconstructed.

3.1 Group-Based Dictionary Construction and Group-Sparse Representation

Our method involves a group of neighbouring patches. Below, we discuss the dictionary construction for this group, followed by our group-sparse representation.

Dictionary Construction for the Central Patch

The dictionary construction for the central patch is similar to that in [7]. The 2D patches $\mathbf{y}_c^U$ and $\mathbf{y}_c^L$ from $I(s)$ and $I(s+1)$, respectively, and their x and y gradients $(\mathbf{F}^x(\mathbf{y}_c^U), \mathbf{F}^y(\mathbf{y}_c^U), \mathbf{F}^x(\mathbf{y}_c^L), \mathbf{F}^y(\mathbf{y}_c^L))$ are used as observations. We search for candidate patches $\mathbf{y}_{p_c}$ in a 3D region in each phase p (other than the current phase), around the same locations as that of $\mathbf{y}_c^U$ and $\mathbf{y}_c^L$. We select the K best patches yielding the K lowermost costs. Observing that $\mathbf{y}_c^U$ and $\mathbf{y}_c^L$ may be dissimilar, we incorporate a balancing condition in the overall cost E_c^d as defined below.

$$E_c^d = E_c^U + E_c^L \quad \text{if} \quad 1/\epsilon < E_c^U/E_c^L < \epsilon, \quad \text{and} \quad E_c^d = \infty \quad \text{otherwise} \tag{2}$$

where the sub-costs E_c^U and E_c^L, involving $\mathbf{y}_c^U$ and $\mathbf{y}_c^L$, respectively, are defined as

$$\begin{aligned} E_c^U &= ||\mathbf{y}_c^U - \mathbf{y}_{p_c}||_2 + \gamma(||\mathbf{F}^x(\mathbf{y}_c^U) - \mathbf{F}^x(\mathbf{y}_{p_c})||_2 + ||\mathbf{F}^y(\mathbf{y}_c^U) - \mathbf{F}^y(\mathbf{y}_{p_c})||_2) \\ E_c^L &= ||\mathbf{y}_c^L - \mathbf{y}_{p_c}||_2 + \gamma(||\mathbf{F}^x(\mathbf{y}_c^L) - \mathbf{F}^x(\mathbf{y}_{p_c})||_2 + ||\mathbf{F}^y(\mathbf{y}_c^L) - \mathbf{F}^y(\mathbf{y}_{p_c})||_2) \end{aligned} \tag{3}$$

where γ is the weighting of the gradient feature cost. Thus, only those patches which yield a low-cost and are similar to both $\mathbf{y}_c^U$ and $\mathbf{y}_c^L$ are included in the dictionary.

A similar *balancing* as in equation (2) is also used in [7], but during the greedy sparse representation step. We follow a non-greedy optimization for sparse representation, where the balancing term yields non-linearities. Hence, we incorporate it during dictionary construction, to resist imbalanced patches from entering the optimization.

Dictionary Construction for Neighbouring Patches

Having defined the dictionary for the central patch (say D_c), we now construct N neighbourhood dictionaries D_1 to D_N using observed patches $\mathbf{y}_1^U, ..., \mathbf{y}_N^U$ and $\mathbf{y}_1^L, ..., \mathbf{y}_N^L$ neighbouring to the central patch $\mathbf{y}_c^U$ and $\mathbf{y}_c^L$, respectively. This proceeds as follows: a) Patches in $D_1, ..., D_N$ are selected such that their spatial relationship with those in D_c is consistent as that of $\mathbf{y}_1^U, ..., \mathbf{y}_N^U$ to $\mathbf{y}_c^U$ (and $\mathbf{y}_1^L, ..., \mathbf{y}_N^L$ to $\mathbf{y}_c^L$). For instance, if $\mathbf{y}_1^U$ is left to $\mathbf{y}_c^U$, then the patches in D_1 are left to those in D_c. This ensures that the spatial correspondence among neighbouring candidate patches is same as that of neighbouring observed patches. b) To induce a constructive smoothness, the candidate patches that contribute to the dictionaries satisfy a soft-similarity condition (equation (4)), so as to avoid the inclusion of unsuitable patches in the dictionaries.

$$\mathbf{y}_{p_n} \in D_n \quad \text{if} \quad E_n^d < \kappa E_c^d \tag{4}$$

where $\mathbf{y}_{p_n}$ is the candidate patch for dictionary D_n, and E_n^d is the corresponding cost for $\mathbf{y}_{p_n}$ (defined similarly to E_c^d, using proper change in subscripts). κ is a threshold parameter. Note that this condition plays a similar (rather, a softer) role as neighbourhood similarity [4], which advocates that neighbours tend to be similar in appearance. Hence, it is fair to assume that neighbouring patches typically tend to have similar costs. Indeed, the condition helps in excluding those dissimilar neighbouring patches, which may not be suitable in group-sparse representation for inducing a constructive smoothness.

For every patch, we keep the number of atoms in all dictionaries (including D_c) equal. Hence, if the above condition is not satisfied by all the neighbouring patches, we do not include the corresponding patch in any dictionary.

Group-Sparse Coefficient Estimation and Reconstruction

We use the constructed dictionaries to solve for the coefficients that minimizes the group-sparse problem of equation (1), rewritten for our case in the left equation in (5):

$$\hat{A} = \arg\min_{\alpha_k} \sum_{k\in\{\{1,\ldots,N\},c\}} ||\tilde{\mathbf{y}}_k - D_k\alpha_k||_2^2 + \lambda||A||_{2,1} \qquad \tilde{\mathbf{y}}_c^O = D_c\hat{\alpha}_c \tag{5}$$

where $\tilde{\mathbf{y}}_k$ is the concatenated observation vector for the k^{th} task defined as: $\tilde{\mathbf{y}}_k = [(\mathbf{y}_k^U)^T,\ \ (\gamma\mathbf{F}^x(\mathbf{y}_k^U))^T,\ \ (\gamma\mathbf{F}^y(\mathbf{y}_k^U))^T,\ \ (\mathbf{y}_k^L)^T,\ \ (\gamma\mathbf{F}^x(\mathbf{y}_k^L))^T,\ \ (\gamma\mathbf{F}^y(\mathbf{y}_k^L))^T]^T$. To maintain dimensional consistency with $\tilde{\mathbf{y}}_k$, the atoms in the dictionary are defined as $[(\tilde{\mathbf{y}}_{p_k})^T,\ \ (\tilde{\mathbf{y}}_{p_k})^T]^T$, where $\tilde{\mathbf{y}}_{p_k} = [(\mathbf{y}_{p_k})^T,\ \ (\gamma\mathbf{F}^x(\mathbf{y}_{p_k}))^T,\ \ (\gamma\mathbf{F}^y(\mathbf{y}_{p_k}))^T]^T$.

The coefficient matrix $\hat{A}$ is estimated via convex optimization of left equation in (5) [9]. Following this, the output vector $\tilde{\mathbf{y}}_c^O$ corresponding to the central patch is reconstructed as in the right equation in (5) using $\hat{\alpha}_c$, the coefficient vector corresponding to the center patch (which is extracted from the matrix $\hat{A}$ and normalized to 1). The top one-sixth of $\tilde{\mathbf{y}}_c^O$ (containing raw intensities) is reshaped and placed into the unknown slice, with overlapping values properly averaged.

Observe that although the coefficients are estimated for all dictionaries, we only reconstruct the central patch, i.e. using only the dictionary D_c and coefficient vector $\hat{\alpha}_c$. It might appear that the group-sparsity is not being exploited during the patch reconstruction. However, we note that group-sparsity regularization guides the coefficient estimation (including $\hat{\alpha}_c$), thus enabling the selection of those dictionary patches from D_c for reconstruction and better inducing the controlled smoothness.

3.2 Two-Stage Strategy

The above approach provides a reasonable reconstruction, but has further scope of improvement. Hence, we employ a second step that improves upon the above reconstruction. We now use only the patches from the reconstructed slice.

The dictionary construction cost for the central patch using the patch $\mathbf{y}_c^R$ in the reconstructed slice, and its gradients $\mathbf{F}^x(\mathbf{y}_c^R)$ and $\mathbf{F}^y(\mathbf{y}_c^R)$, is similar to equation (3) as

$$E_c^d = ||\mathbf{y}_c^R - \mathbf{y}_{p_c}||_2 + \gamma(||\mathbf{F}^x(\mathbf{y}_c^R) - \mathbf{F}^x(\mathbf{y}_{p_c})||_2 + ||\mathbf{F}^y(\mathbf{y}_c^R) - \mathbf{F}^y(\mathbf{y}_{p_c})||_2) \tag{6}$$

Note that here we do not require any balancing condition. The neighbourhood dictionaries are also constructed similarly as before, except with $\mathbf{y}_1^R, ..., \mathbf{y}_N^R$ as the patches neighbouring to $\mathbf{y}_c^R$. Finally, sparse representation and reconstruction is carried out as

$$\hat{A} = \arg\min_{\alpha_k} \sum_{k \in \{\{1,...,N\},c\}} ||\tilde{\mathbf{y}}_k^R - D_k \alpha_k||_2^2 + \lambda ||A||_{2,1} \quad \text{and} \quad \tilde{\mathbf{y}}_c^F = D_c \hat{\alpha}_c \quad (7)$$

where $\tilde{\mathbf{y}}_k^R = [(\mathbf{y}_k^R)^T, \ (\gamma \mathbf{F}^x(\mathbf{y}_k^R))^T, \ (\gamma \mathbf{F}^y(\mathbf{y}_k^R))^T]^T$, and the dictionary atoms for the selected patches are denoted as: $\tilde{\mathbf{y}}_{p_k} = [(\mathbf{y}_{p_k})^T, \ (\gamma \mathbf{F}^x(\mathbf{y}_{p_k}))^T, \ (\gamma \mathbf{F}^y(\mathbf{y}_{p_k}))^T]^T$. $\tilde{\mathbf{y}}_c^F$ is the output vector which is used in reconstructing the final output slice.

The need for the second step indicates that the group-sparsity method works better given a 'good' estimate from the first step. Hence, inspired by [7], a larger scale is chosen in the first step, which yields smooth but structurally more correct results. We then choose a smaller scale in the second step for better structure localization.

While the method in [7] also reduces scale over iterations, the successive outputs are averaged with a large weight to the output at the highest scale. Our second stage output is constructed strictly at a lower scale, thus yielding better structural enhancement and still maintaining robustness to noise and artifacts due to group-sparsity regularization.

4 Experimental Results

Our experiments involve the DIR-Lab lung data [12], containing 10 cases, each with 10 respiratory phases. The in-plane resolution is 1 mm and the superior-inferior resolution is 2.5 mm. We further subsample this data by removing alternate slices, so as to compare our estimated intermediate slice with the true slice. We provide qualitative and quantitative results, and also compare with the method in [7], a related sparse-representation-based approach. As our work is about resolution enhancement, we select a range of slices with a good amount of vessel structures. The quantitative metrics are computed around vessel regions extracted using the method in [13]. Such an error computation emphasizes structural enhancement, better localization and reduction in artifacts, which is central to SR and an indicator of improvement in spatial resolution. Due to space constraints, we do not provide visual results for interpolation methods.

We use the MALSAR package [14] for group-sparse coefficient computation. Our approach involves parameters γ, ϵ, κ and λ which, for the results below, are as follows: $\gamma = 0.2$, $\epsilon = 1.1$, κ is chosen so that dictionary contains 300 - 400 patches, and $\lambda = 40000$ (sparsity of $\sim 30\%$). The patch-size in the first step is 32×32 and in the second is 16×16. The 3D search region is $11 \times 11 \times 11$, and we employ $N = 4$ neighbors to the center patch. It is worth noting that our parameter range for best result is narrow, which simplifies tuning, in general.

4.1 Visual Qualitative Results

We first demonstrate some qualitative slice reconstructions in Fig. 1, which shows some typical outputs. The first to third columns depict reconstruction obtained using the approach in [7], our output, and ground truth, respectively. It can be observed that vessel structures in our outputs are better localized and more accurate than those in [7].

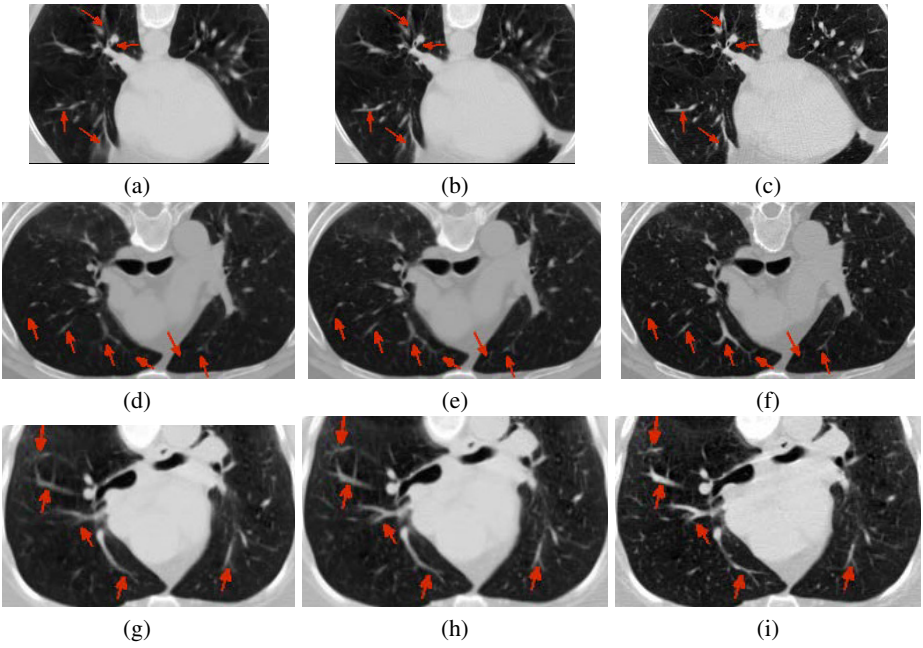


Fig. 1. Examples for reconstruction of some slices: (a,d,g) Outputs using the approach in [7]. (b,e,h) Outputs using the proposed method. (c,f,i) Ground-truth.

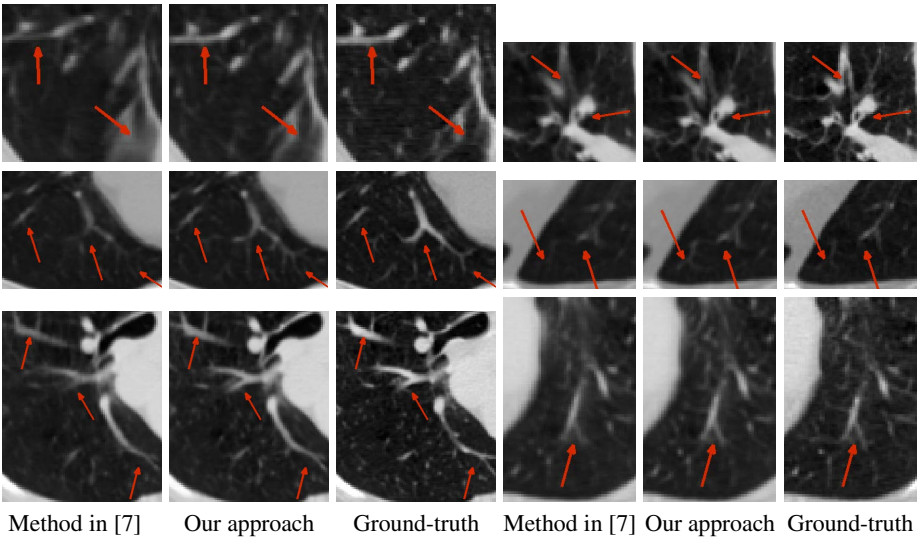


Fig. 2. Close-up views: Each row depicts the close-up views of two regions from slices shown in the corresponding row of Fig. 1. Column pairs $\{1, 4\}, \{2, 5\}, \{3, 6\}$ correspond to the results by the approach in [7], proposed approach and ground-truth, respectively.

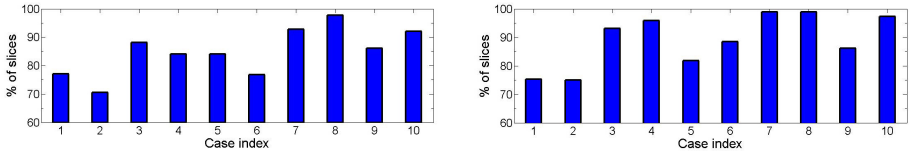
Some of the improvements are marked with red arrows. We zoom into some regions of Fig. 1, and show the close-up views in Fig. 2, where one can further appreciate our method achieving more plausible and better structure reconstruction. While the results are somewhat smooth than ground-truth (which is due to patch averaging), they clearly indicate improved spatial resolution of vessel structures over the state-of-the-art [7].

4.2 Quantitative Results

We next provide quantitative results averaged over complete sets of data used in our experiments. Table 1 provides case-wise root-mean-square (RMS) and structural similarity (SSIM) metrics for bi-cubic interpolation, the method in [7] and our approach. Observe that our approach shows a clear improvement with respect to both metrics for all 10 cases (note that such an order of improvement is common in contemporary super-resolution (e.g. [3,15])). The RMS results support our claims for overall mitigation of the noise-structure trade-off, whereas the SSIM results emphasize better structure enhancement. In Fig. 3 we show the percentage of slices across cases, for which our approach favourably compares with that in [7]. Clearly, our approach better reconstructs a vast majority of slices, which highlights that the group-sparsity-based smoothing indeed has a constructive effect. Overall, the RMS improvement is over 85% of the slices and the SSIM improvement is over 89% of the slices.

Table 1. Average RMS and SSIM for 10 cases

Case (No. of Slices)	RMS: Bicubic	RMS: [7]	RMS: Proposed	SSIM: Bicubic	SSIM: [7]	SSIM: proposed
Case 1 (170)	28.97	18.56	17.96	0.5492	0.7069	0.7127
Case 2 (200)	27.83	16.74	16.32	0.5777	0.7445	0.7511
Case 3 (160)	26.82	15.86	15.15	0.5895	0.7589	0.7734
Case 4 (170)	26.95	17.10	16.36	0.5869	0.7372	0.7550
Case 5 (170)	29.00	19.23	18.24	0.5671	0.7184	0.7329
Case 6 (190)	24.81	17.75	17.25	0.5151	0.6255	0.6421
Case 7 (180)	28.12	18.94	17.94	0.5181	0.6622	0.6856
Case 8 (180)	37.38	26.72	24.70	0.4913	0.6137	0.6462
Case 9 (130)	25.34	16.63	16.04	0.5558	0.6980	0.7123
Case 10 (190)	37.10	25.40	23.74	0.4820	0.6148	0.6461
Average (Total: 1740)	29.23	19.29	18.37	0.5433	0.6860	0.7057



(a) (b)

Fig. 3. % of slices on which our approach performs favourably over [7] for (a) RMS, (b) SSIM

5 Conclusion

We proposed a super-resolution method for 4D-CT, where we discussed the regularizing role of group-sparsity and employed it via careful dictionary construction based on patch neighbourhood and similarity. Our results justify the potential of group-sparsity for noise robustness and structure enhancement. In the future, we aim to explore other group-sparsity frameworks and carry out exhaustive parameter analysis.

References

1. Rit, S., Sarrut, D., Desbat, L.: Comparison of analytic and algebraic methods for motion-compensated cone-beam CT reconstruction of the thorax. IEEE Trans. on Medical Imaging 28(10), 1513–1525 (2009)
2. Bose, N., Ahuja, N.: Superresolution and noise filtering using moving least squares. IEEE Trans. on Image Processing 15(8), 2239–2248 (2006)
3. Yang, J., Wright, J., Huang, T., Ma, Y.: Image super-Resolution via sparse representation. IEEE Trans. on Image Processing 19(11), 2861–2873 (2010)
4. Freeman, W., Jones, T., Pasztor, E.: Example-based super-resolution. IEEE Computer Graphics and Applications 22(2), 56–65 (2002)
5. Goshtasby, A., Turner, D., Ackerman, L.: Matching of tomographic slices for interpolation. IEEE Trans. on Medical Imaging 11(4), 507–516 (1992)
6. Wu, G., Lian, J., Shen, D.: Improving image-guided radiation therapy of lung cancer by reconstructing 4D-CT from a single free-breathing 3D-CT on the treatment day. Med. Phys. 39(12), 7694–7709 (2012)
7. Zhang, Y., Wu, G., Yap, P., Feng, Q., Lian, J., Chen, W., Shen, D.: Hierarchical patch-based sparse representation - A new approach for resolution enhancement of 4D-CT lung data. IEEE Trans. on Medical Imaging 31(11), 1993–2005 (2012)
8. Zhang, Y., Wu, G., Yap, P.-T., Feng, Q., Lian, J., Chen, W., Shen, D.: Non-local means resolution enhancement of lung 4D-CT data. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) MICCAI 2012, Part I. LNCS, vol. 7510, pp. 214–222. Springer, Heidelberg (2012)
9. Liu, J., Ji, S., Ye, J.: Multi-task feature learning via efficient $l_{2,1}$-norm minimization. In: Conference on Uncertainty in Artificial Intelligence (UAI 2009), pp. 339–348 (2009)
10. Mairal, J., Bach, F., Ponce, J., Sapiro, G., Zisserman, A.: Non-local sparse models for image restoration. In: International Conference on Computer Vision (ICCV 2009), pp. 2272–2279 (2009)
11. Yang, C.-Y., Huang, J.-B., Yang, M.-H.: Exploiting self-similarities for single frame super-resolution. In: Kimmel, R., Klette, R., Sugimoto, A. (eds.) ACCV 2010, Part III. LNCS, vol. 6494, pp. 497–510. Springer, Heidelberg (2011)
12. Castillo, R., Castillo, E., Guerra, R., Johnson, V., McPhail, T., Garg, A., Guerrero, T.: A framework for evaluation of deformable image registration spatial accuracy using large landmark point sets. Phys. Med. Biol. 54, 1849–1870 (2009)
13. Heng, J., Chen, C., Cole, E., Pisano, E., Shen, D.: Automated delineation of calcified vessels in mammography by tracking with uncertainty and graphical linking techniques. IEEE Trans. on Medical Imaging 31(11), 2143–2155 (2012)
14. Zhou, J., Chen, J., Ye, J.: MALSAR: Multi-tAsk Learning via StructurAl Regularization. Arizona State University (2012)
15. Takeda, H., Milanfar, P., Protter, M., Elad, M.: Super-Resolution Without Explicit Subpixel Motion Estimation. IEEE Trans. on Image Processing 18(9), 1958–1975 (2009)