Resolution enhancement of lung 4D-CT via group-sparsity

Arnav Bhavsar[a] and Guorong Wu[a]
Department of Radiology and BRIC, University of North Carolina, Chapel Hill, North Carolina 27599

Jun Lian[b]
Department of Radiation Oncology, University of North Carolina, Chapel Hill, North Carolina 27599

Dinggang Shen[a]
Department of Radiology and BRIC, University of North Carolina, Chapel Hill, North Carolina 27599

(Received 5 July 2013; revised 2 September 2013; accepted for publication 20 October 2013; published 15 November 2013)

Purpose: 4D-CT typically delivers more accurate information about anatomical structures in the lung, over 3D-CT, due to its ability to capture visual information of the lung motion across different respiratory phases. This helps to better determine the dose during radiation therapy for lung cancer. However, a critical concern with 4D-CT that substantially compromises this advantage is the low superior-inferior resolution due to less number of acquired slices, in order to control the CT radiation dose. To address this limitation, the authors propose an approach to reconstruct missing intermediate slices, so as to improve the superior-inferior resolution.
Methods: In this method the authors exploit the observation that sampling information across respiratory phases in 4D-CT can be complimentary due to lung motion. The authors' approach uses this locally complimentary information across phases in a patch-based sparse-representation framework. Moreover, unlike some recent approaches that treat local patches independently, the authors' approach employs the group-sparsity framework that imposes neighborhood and similarity constraints between patches. This helps in mitigating the trade-off between noise robustness and structure preservation, which is an important consideration in resolution enhancement. The authors discuss the regularizing ability of group-sparsity, which helps in reducing the effect of noise and enables better structural localization and enhancement.
Results: The authors perform extensive experiments on the publicly available DIR-Lab Lung 4D-CT dataset [R. Castillo, E. Castillo, R. Guerra, V. Johnson, T. McPhail, A. Garg, and T. Guerrero, "A framework for evaluation of deformable image registration spatial accuracy using large landmark point sets," Phys. Med. Biol. **54**, 1849–1870 (2009)]. First, the authors carry out empirical parametric analysis of some important parameters in their approach. The authors then demonstrate, qualitatively as well as quantitatively, the ability of their approach to achieve more accurate and better localized results over bicubic interpolation as well as a related state-of-the-art approach. The authors also show results on some datasets with tumor, to further emphasize the clinical importance of their method.
Conclusions: The authors have proposed to improve the superior-inferior resolution of 4D-CT by estimating intermediate slices. The authors' approach exploits neighboring constraints in the group-sparsity framework, toward the goal of achieving better localization and noise robustness. The authors' results are encouraging, and positively demonstrate the role of group-sparsity for 4D-CT resolution enhancement. © *2013 American Association of Physicists in Medicine*. [http://dx.doi.org/10.1118/1.4829501]

Key words: group-sparsity, resolution enhancement, 4D-CT, radiation therapy, lung cancer

1. INTRODUCTION

Lung cancer is a growing concern, being the cause for majority of cancer-related deaths. Radiation therapy is widely used as an effective treatment for lung cancer, where x-ray radiation is used to kill the tumor. In this regard, accurate localization of tumor and other lung structures is important to plan the therapy so that the radiation more controls the tumor and less affects the surrounding lung structures. However, due to the respiratory lung motion, acquiring such accurate structural information is difficult in traditional free-breathing 3D-CT which captures a single volume over the complete respiratory cycle, and hence suffers from motion artifacts due to the moving structures in the lung.

To address this concern, in recent years, 4D-CT imaging has been developed which acquires a volume for each respiratory phase. This provides more accurate estimate of the lung motion across respiratory phases of the breathing cycle and much better localization of the moving structures in the lung. This, in turn, helps in better planning of the radiation dose by taking the respiratory motion and structural information into account.

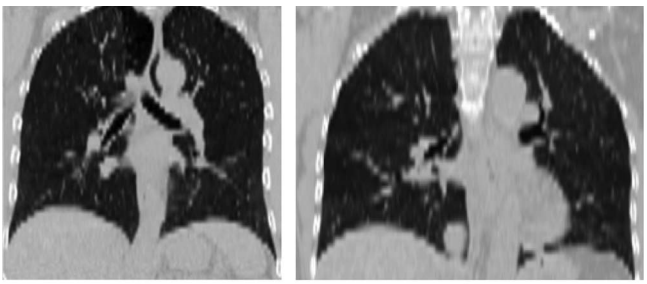
The growing importance of 4D-CT is also supported by various recent works on 4D-CT reconstruction.[1,2] Such

approaches consider the sorting and registration of the different slices in volumes corresponding to different respiratory phases. Moreover, various clinical works have also been reported, which highlights the use of 4D-CT in planning the radiation therapy.[3,4] Thus, the usefulness of 4D-CT is clearly being appreciated in the medical imaging and radiological community.

Having stated the importance of 4D-CT, it must be asserted that there is an important concern that compromises the advantages of 4D-CT. In order to control the radiation dose during acquisition of 4D-CT imagery, usually a reduced number of slices are acquired. This reduction in the number of slices adversely affects the superior-inferior resolution. For instance, in the publicly available dataset,[5] the in-plane resolution is of the order of 1 mm but the superior-inferior resolution is as low as 2.5 mm. Such a low resolution results in a low image-quality.[6] The low-resolution 4D-CT imagery suffers from false apparent vessel discontinuities, shape distortions, etc., which introduce errors in the correct localization and shape assessment of tumor and vessel structures. Thus, the reduction in resolution compromises the full potential of 4D-CT for providing accurate motion and localization, and the problem of enhancing resolution of 4D-CT is an important one and needs to be addressed.

Resolution enhancement in 4D-CT implies estimating the intermediate slices which were not physically acquired. A common approach to approximate such a reconstruction is by using off-the-shelf interpolation approaches. Specifically, classic interpolation methods use a small set of immediately neighboring slices from the same volume in which the intermediate slices are to be reconstructed. However, as depicted in Fig. 1 for two example coronal views, such interpolation schemes also yield the above mentioned errors and artifacts, (such as inaccurate, jagged, and discontinuous vessels structures) and clearly do not serve the purpose of enhancing the resolution.

Indeed, more sophisticated approach for resolution enhancement, often known as super-resolution (SR) methods, look for the lost sampling information in different instances of data.[7] In a similar spirit, we exploit the observation that due to lung motion, slices at the similar positions would often sample different content of lung in different respiratory phases. Thus, the sampling information can be complementary across respiratory phases. Hence, information lost in some respiratory phase can be found in others, which, in turn, can be used to estimate the unknown intermediate slices.



(a) (b)

FIG. 1. (a,b) Two examples from the coronal view of 4D-CT where the intermediate axial slices are constructed using conventional bicubic-interpolation. The image artifacts (zigzags) and the motion artifacts (discontinuous vessels), due to inaccurately estimated axial slices, are obvious.

Since the lung motion is deformable, such a search for interphase information would be more useful if made locally, i.e., one must look for local patches across phases to reconstruct the intermediate slice in certain phase. Moreover, as there can be numerous possible selections of candidate patches that can agree toward the required reconstruction, it is advisable to select only a few patches that best agree with some reconstruction criteria (e.g., some similarity cost). Another important consideration, especially for resolution enhancement, is that of reducing the effect of noise while still preserving the high frequency structures;[8] a challenging task as it involves a trade-off as noise is also typically a high-frequency phenomenon.

From a clinical point-of-view, enhancing the accuracy and localization of anatomical structures is important for a radiologist to correctly assess the shape and extent of tumor and nearby anatomical structures. At the same time, it is also crucial that the effect of noise and presence of artifacts is minimized in the reconstructed slices. Hence, the practical importance of mitigating the "noise-structure trade-off" can be clearly appreciated.

These considerations lead us to employ a patch-based group-sparse representation framework for the task of resolution enhancement. More specifically, we reconstruct the intermediate slices patchwise, where for each patch we reconstruct multiple dictionaries, each containing patches from local neighborhoods from the observed slices across phases which potentially contain the required sampling information. The construction of these neighborhood dictionaries is an important step which supports the following process of group-sparse representation. The sparsity part encourages the selection of few patches out of the dictionary that best satisfies some criterion. The grouping part, which uses the multiple neighborhood dictionaries, constrains the sparsity in such a manner so as to induce smoothness during selection of patches that help mitigate the effect of noise.

1.A. Relation to previous work

As mentioned above, the philosophy of using of interphase information for resolution enhancement, as a result of the lung-motion, is on similar lines as that in super-resolution approaches in computer vision,[7] which also search for lost sampling information across different sources or instances of data. Typically, super-resolution methods fall into one of two categories. The motion-based super-resolution methods use multiple observations with sub-pixel relative motion,[7,9] where the high-resolution (HR) information is embedded in the observations due to the subpixel shifts. They then use the registration information to integrate information from such shifted images to yield an high-resolution image. On the other hand, learning-based or example-based super-resolution methods use a large dictionary typically containing thousands of local high and low resolution patch pairs from a separate off-line

high-resolution image dataset.[10,11] The HR patches which are paired with those LR patches from the dictionary, that match the local LR patches in the observation, are used to reconstruct the HR image. Indeed, some of the more successful learning-based methods also employ a sparse-representation framework for estimating the HR image.[10] Such motion-based and learning-based frameworks have also been employed for super-resolution in the medical image domains, especially for MR images.[12,13]

However, looking more closely, the 4D-CT resolution enhancement problem that we consider is quite distinct from the above discussed SR approaches in terms of data usage and availability. This is because, unlike motion-based methods, it is difficult to rely on accurate registration in 4D-CT because of the poor superior-inferior resolution.[14,15] Also, unlike the learning-based approaches, the availability of large off-line HR datasets for 4D-CT which are used for constructing dictionaries is infeasible. Hence, such traditional perspectives may not be directly applicable for 4D-CT resolution enhancement. Indeed, our approach can be viewed as the one which integrating ideas from both domains, i.e., in case of 4D-CT resolution enhancement the unique sampling information is induced by motion but is used in a learning-based framework.

Another important difference between the traditional super-resolution methods and the approach that we propose in this work is in the manner of incorporating smoothness constraints. In addition to considering similarity of local information from different sources with the observations, most super-resolution approaches follow different regularization strategies to induce smoothness for noise robustness. Global regularization frameworks such as Tikhonov prior,[16] total variation,[17] Markov random fields,[11] etc. are common to both motion-based and learning-based approaches. On the other hand, our approach is essentially local, based on a patchwise reconstruction, and also seeks the advantage of the sparsity framework, where, however, it is not straightforward to include smoothness terms among neighboring patches. Hence to incorporate smoothness, as we will discuss, we harness the regularizing potential of group-sparsity that constrains the sparse representation for spatially neighboring and similar patches. This also allows to retain the successful sparse-representation framework while also providing a smoothness ability.

To our knowledge, an application of group-sparsity toward inducing smoothness is only reported in Refs. 18 and 19. The work reported in Refs. 18 and 19 use group-sparsity for image denoising and super-resolution, respectively. However, the dictionary consideration for group-sparsity in these works is very different from ours. Nevertheless, these works also support the role of group-sparsity for regularization, which we harness for our case of 4D-CT resolution enhancement.

Indeed, the problem of resolution enhancement for 4D-CT is itself a very recent exploration with only a couple of reported works.[15,20] Both these approaches also follow a interphase patch-based reconstruction of intermediate slices, similar to ours. In fact, our approach is closely related to that in Ref. 15, which is also based on sparse representation. However, the method in Ref. 15 does not employ any neighborhood or similarity constraints between neighboring patches, which essentially renders it devoid of regularization of smoothness capabilities. Perhaps, this necessitates the use of relatively large scale of patches for better performance, so as to resist the effect of noise and other artifacts. However, a large scale results in over-smoothing, due to greater averaging of voxels. On the other hand, a smaller scale to improve localization could result in an error-prone noisy reconstruction.

In this regard, our approach which is established within a group-sparsity framework involves neighborhood patches to induce smoothness constraints and better noise robustness. This enables us to constructively use a smaller patch scale (as usually advocated in learning-based methods[10,11]) that helps in enhancing local structures without compromising image quality (e.g., in terms of error metrics). This helps us in mitigating the noise-structure trade-off.

Thus, the contributions of this work over the above discussed related works are that (1) it further advances the recent and clinically important area of 4D-CT resolution enhancement, via an approach that is inspired by traditional super-resolution methods, but is established within a different framework than such methods. (2) It explores the role of group-sparsity in image reconstruction, which is also little reported, and which could be adapted to other image analysis areas. (3) Unlike the existing work on 4D-CT resolution enhancement, our work exploits neighborhood constraints using a novel construction of dictionaries, and uses these in the group-sparse framework for better structural localization.

The paper is organized as follows: In Sec. 2 we provide a discussion on group-sparsity and an indication of how we employ it. Section 3 elaborates our overall methodology. We discuss our experiments and provide qualitative and quantitative results in Sec. 4 and conclude in Sec. 5.

2. GROUP-SPARSITY FOR REGULARIZATION

Before venturing into the elaboration of our approach, we first discuss about group-sparsity and its role in regularization/smoothness. As, to our knowledge, this is one of the early works that exploits group-sparsity for incorporating smoothness in image reconstruction,[18,19] the aim of such a discussion is to indicate the rationale behind some important aspects of our approach such as the use of neighboring patches and dictionary construction.

The traditional framework of sparse representation treats a relatively large set of patches (known as dictionary), as a basis to reconstruct other patches, which presumably lie in the same space as the dictionary patches. The premise for sparse representation is that if the dictionary is representative enough, then ideally any patch can be expressed as a linear combination of *few* of the patches in the dictionary. Hence, the sparse-representation problem is that of estimating which dictionary patches are used in such a linear combination, and their corresponding weights.

Formally, a sparse representation problem can be defined as[10]

$$\hat{\alpha} = \arg\min_{\alpha} ||\mathbf{y} - D\alpha||_2^2 + \lambda||\alpha||_1, \quad (1)$$

where $\mathbf{y}$ denotes a M_1-dimensional observation vector, and D is the $M_1 \times M_2$ dictionary matrix (columns corresponding to M_2 different atoms). The M_2-dimensional coefficient vector α describes the linear combination of atoms in D that best matches $\mathbf{y}$. The l_1-norm $||\alpha||_1$ enforces sparsity on the components of α, so that only few dictionary atoms are selected. For patch-based image reconstruction, the output patch is reconstructed as a linear combination of the dictionary patches corresponding to non-zero estimated coefficients, where the coefficients serve as the weights in the linear combination.

While the sparsity constraint is effective for many problems, in many applications, multiple factors (also referred to as "tasks") can be related to the estimation problem. In such cases, the notion of group-sparsity[21] is quite useful to consider these tasks in a single framework. Thus, group-sparse representation is essentially an extension to the sparse representation problem of Eq. (1). More specifically, the group-sparse problem involving K tasks can be expressed as

$$\hat{A} = \arg\min_{\alpha_l} \sum_{l=1}^{K} ||\mathbf{y}_l - D_l\alpha_l||_2^2 + \lambda||A||_{2,1}. \quad (2)$$

Here, $\mathbf{y}_l$ is the observation and D_l is the dictionary corresponding to the lth task. $A = [\alpha_1, \ldots, \alpha_l, \ldots, \alpha_K]$ is an $M_2 \times K$ matrix whose columns correspond to the coefficient vectors related to the K different tasks. $||A||_{2,1}$ is referred to as the $l_{2,1}$-norm over the sparse coefficients and defined as

$$||A||_{2,1} = \sum_{q=1}^{M_2} ||\mathbf{a}^q||_2, \quad (3)$$

where $\mathbf{a}^q$ denotes the qth row of A and the $l_{2,1}$-norm $||A||_{2,1}$ is an l_1-norm over the l_2-norms of the rows of A.[21,22,23]

The $l_{2,1}$-norm $||A||_{2,1}$ essentially enforces coefficients related to different tasks to follow a similar sparsity structure, i.e., the location of the non-zero elements is same across all columns of A (although their values may differ). For instance, as shown in Fig. 2(a), the coefficient vectors for related tasks, when estimated independently via traditional sparse representation acting individually on each task, can be quite different from each other. However, if these tasks are used in a group-sparse framework, then the estimated coefficient vectors share common locations of non-zero coefficients [Fig. 2(b)].

The intuitive idea is that, related tasks should share similar sparsity structure. That is, they should employ correspondingly related dictionary atoms for their description. The notion of *relatedness* is, however, subjective and would differ with the problem at hand. Also, the usefulness of constraining the sparsity structure across dictionaries would also be realized only if the construction of these dictionaries support such relatedness.

In this work, we aim to consider group-sparsity in the context of exploiting its potential for regularization or local smoothness. We define a neighborhood as a set of overlapping patches around a central patch, and the "task" in our case involves patches from a particular central position or its neighboring positions. The multiple dictionaries are constructed with patches from selected candidate neighborhoods

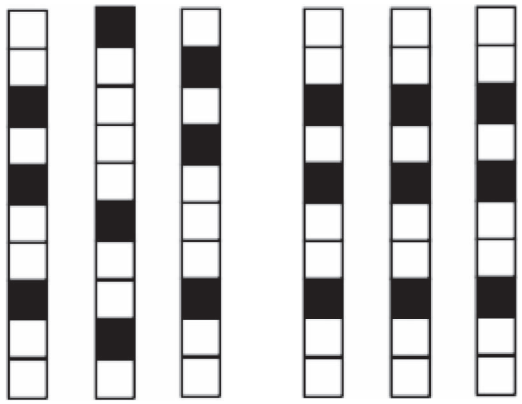


FIG. 2. Depiction of estimated weights across multiple tasks for (a) sparse representation and (b) group-sparse representation.

from slices across different respiratory phases (elaborated in Sec. 3). A dictionary corresponding to each observed neighborhood patch contains the corresponding candidate neighborhood patches from the other phases. While constructing these dictionaries, the selection of candidate neighborhood patches is also based on some similarity condition (we discuss these points in more detail in Sec. 3 elaborating our methodology). Hence, in our context, *relatedness* implies spatial neighborhood and similarity.

When computing a group-sparse representation with dictionaries constructed in such a manner, the selected dictionary patches (with nonzero coefficients) for reconstructing the output patch, would be constrained so as to satisfy a greater closeness with their neighbors. This would, in turn, result in better smoothness and hence noise robustness. Moreover, the above mentioned similarity conditions also ensure that the smoothness does not sacrifice the salient structures. Thus, group-sparsity plays a regularizing role and helps in mitigating the noise-structure trade-off. We discuss more details in subsequent sections where we elaborate our methodology.

3. METHOD

We now formalize our group-sparsity based 4D-CT resolution enhancement approach. Given 4D-CT data $I = \{I_i(s)|i = 1, \ldots, P; s = 1, \ldots, S\}$ (with P phases and S slices), we estimate a slice between $I_i(s)$ and $I_i(s+1)$ for the ith phase in a patchwise manner.[15] The slices $I_i(s)$ and $I_i(s+1)$, which are the upper and lower slices to the slice to be reconstructed. We use the patches from these slices to form the observation vectors.

As indicated above, our method involves the central patch to be reconstructed and its immediate neighbors, where the dictionaries are constructed for each patch in this group. As the central patch plays the primary role in reconstruction, the construction of dictionaries containing the neighboring patches is based on the dictionary for the central patch. Such a strategy also has a secondary advantage of reducing the computation since the search for candidate neighborhoods can be otherwise combinatorial. In addition, we also perform a PCA

on the dictionary atoms, before computing the group-sparse representation. This is motivated from the observation that the top fraction of principal components is typically more robust to noise,[24] which would further help us in mitigating the noise-structure trade-off.

Our method involves two stages, where the second stage refines the reconstructed slice [say, $I_i^R(s+0.5)$] from the first stage. While the first stage uses $I_i(s)$ and $I_i(s+1)$ as the observed slices (which are immediately upper and lower to the slice to be reconstructed), the second stage only uses reconstructed slice $I^R(s+0.5)$ as the observed slice. For both the stages, our method consists of (1) dictionary construction for the central patch, (2) dictionary construction for the neighboring patches, and (3) group-sparse representation. We elaborate these parts for both the stages. In what follows, patches are used as lexicographic vectors and, for ease of notation, we do not use the subscript denoting the phase in which the slice is reconstructed.

3.A. Dictionary construction for the central patch

The dictionary construction for the central patch is similar to that in Ref. 15. We construct a separate dictionary for each patch $\mathbf{y}_i$ to be reconstructed. The 2D center patches $\mathbf{y}_c^U$ and $\mathbf{y}_c^L$ in the observed slices (center patches in slice "s" and "s + 1" in Fig. 3), and their respective 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 dictionary patches in a 3D region in all phases other than the current phase, around the same voxel locations as that of $\mathbf{y}_c^U$ and $\mathbf{y}_c^L$. We select the K best patches yielding the K lowermost costs defined below. Observing that $\mathbf{y}_c^U$ and $\mathbf{y}_c^L$ may be dissimilar to each other in many cases (which can result in dictionary patches being biased toward $\mathbf{y}_c^U$ or $\mathbf{y}_c^L$), we incorporate a balancing condition in the overall cost E_c^d as defined below

$$E_c^d = E_c^U + E_c^L \quad \text{if} \quad 1/\epsilon < E_c^U/E_c^L < \epsilon,$$
$$E_c^d = \infty \quad \text{otherwise}, \tag{4}$$

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

$$\begin{aligned} E_c^U &= \left\|\mathbf{y}_c^U - \mathbf{y}_{p_c}\right\|_2 + \gamma\left(\left\|\mathbf{F}^x\left(\mathbf{y}_c^U\right) - \mathbf{F}^x(\mathbf{y}_{p_c})\right\|_2\right. \\ &\quad \left. + \left\|\mathbf{F}^y\left(\mathbf{y}_c^U\right) - \mathbf{F}^y(\mathbf{y}_{p_c})\right\|_2\right), \\ E_c^L &= \left\|\mathbf{y}_c^L - \mathbf{y}_{p_c}\right\|_2 + \gamma\left(\left\|\mathbf{F}^x\left(\mathbf{y}_c^L\right) - \mathbf{F}^x(\mathbf{y}_{p_c})\right\|_2\right. \\ &\quad \left. + \left\|\mathbf{F}^y\left(\mathbf{y}_c^L\right) - \mathbf{F}^y(\mathbf{y}_{p_c})\right\|_2\right). \end{aligned} \tag{5}$$

Here, $\mathbf{y}_{p_c}$ is a candidate patch from phase p and $\mathbf{F}^x(\mathbf{y}_{p_c})$ and $\mathbf{F}^y(\mathbf{y}_{p_c})$ are its gradient features. 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.

Note that a similar *balancing* as in Eq. (4) is used in Ref. 15. It is incorporated during the sparse learning phase that is based on a greedy strategy. We follow a nongreedy optimization for our sparse learning, where such a balancing term results in nonlinearities. Instead, we incorporate the balancing during dictionary construction, which reduces the

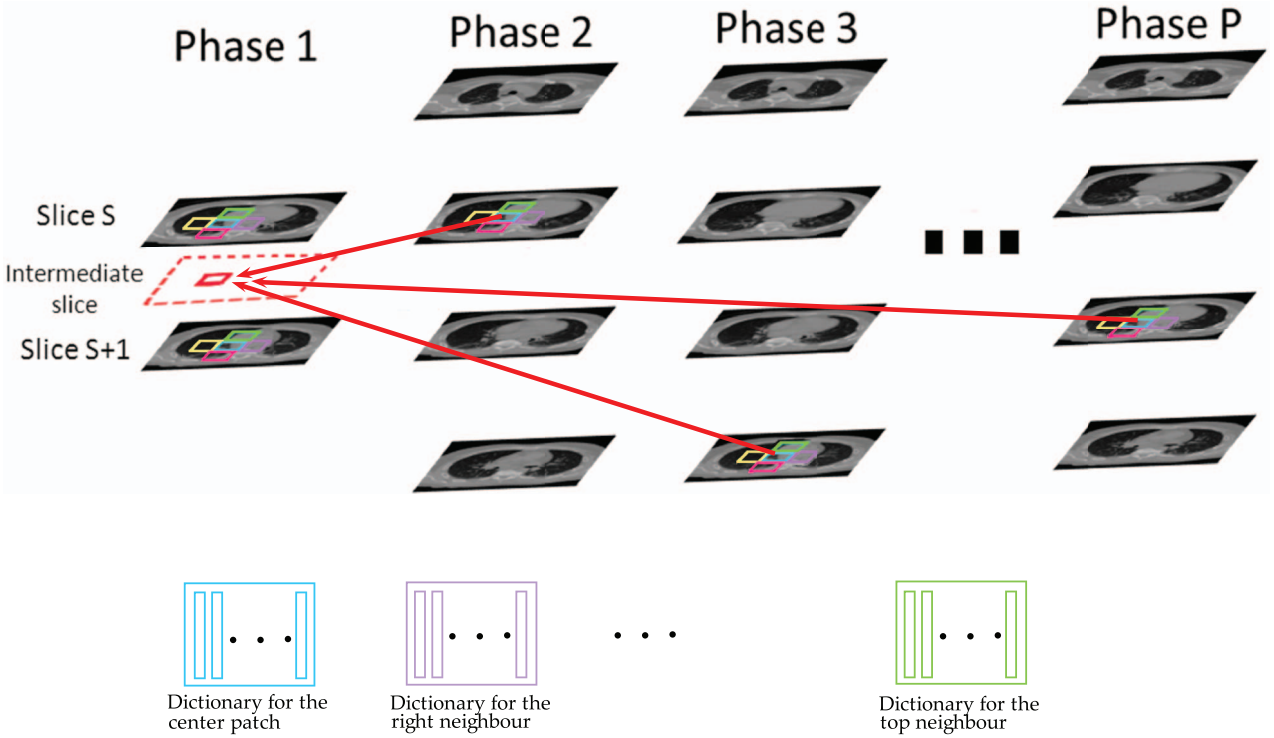


FIG. 3. Dictionary construction and patch reconstruction: In the above example figure, the patches in phases 2 to P, denote the center and neighboring patch candidates to be added in their respective dictionaries (also shown in the bottom row). For example, the center patch candidates are added to the first dictionary in the bottom row and the corresponding top neighbors are added to the last dictionary in the bottom row. The corresponding observed patches are shown in the observed slices of the reference phase (phase 1, here). The arrows show that the selected center candidate patches with non-zero coefficients are linearly combined to reconstruct the unknown center patch.

possibility of imbalanced patches entering the group-sparsity optimization process.

The second stage also uses similar cost formulations as the first stage, except that we only use the patches from the reconstructed slice in the first. 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 defined analogous to Eq. (5) as

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

Note that here we do not require any balancing condition as we have only one patch as the observation, unlike the two patches in the first stage.

3.B. Dictionary construction for neighboring patches

Having defined the dictionary (say D_c) for the central patch, we now define N neighborhood dictionaries D_1 to D_N for the N patches neighboring to the central patch. To construct these dictionaries, we use as observations, the patches $\mathbf{y}_1^U, \ldots, \mathbf{y}_N^U$ and $\mathbf{y}_1^L, \ldots, \mathbf{y}_N^L$, which are neighboring to the central patch $\mathbf{y}_c^U$ and $\mathbf{y}_c^L$, respectively (the top, left, right, and bottom patches in the observed slices of the reference phase 1 in Fig. 3), and their corresponding gradients.

The construction of neighborhood dictionaries proceeds as follows: (a) To incorporate neighborhood constraints, patches in $D_1, \ldots, D_N$ are selected such that their spatial relationship with those in D_c is consistent with that of $\mathbf{y}_1^U, \ldots, \mathbf{y}_N^U$ to $\mathbf{y}_c^U$ (and $\mathbf{y}_1^L, \ldots, \mathbf{y}_N^L$ to $\mathbf{y}_c^L$). For instance, if $\mathbf{y}_1^U$ is left adjacent to $\mathbf{y}_c^U$, then the patches in D_1 should also be left adjacent to those in D_c (as also shown in Fig. 3 and explained in its caption). This ensures that only neighboring candidate patches from the other phases would be employed in the reconstruction of corresponding neighboring output patches. (b) Furthermore, to induce a constructive smoothness among neighboring patches, the candidate neighborhoods that should contribute to the dictionaries should satisfy certain similarity conditions, so as to avoid the inclusion of unsuitable neighborhood patches in the dictionaries. Only those neighborhood patches are selected which satisfy a soft-similarity (as defined below) with the center patch.

The similarity with the observed patches $\mathbf{y}_1^U, \ldots, \mathbf{y}_N^U$ and $\mathbf{y}_1^L, \ldots, \mathbf{y}_N^L$ is defined similarly as that for the center patch [Eqs. (4) and (5)], except using the observed neighboring patches $\mathbf{y}_n^U$ and $\mathbf{y}_n^L$ ($n = 1 \ldots N$), and candidate neighboring patches $\mathbf{y}_{p_n}$ from another phase. This cost is expressed as

$$E_n^d = E_n^U + E_n^L \quad \text{if} \quad 1/\epsilon < E_n^U/E_n^L < \epsilon,$$
$$E_n^d = \infty \quad \text{otherwise} \quad (7)$$

where

$$E_n^U = \left\|\mathbf{y}_n^U - \mathbf{y}_{p_n}\right\|_2^2 + \gamma\left(\left\|\mathbf{F}^x\left(\mathbf{y}_n^U\right) - \mathbf{F}^x(\mathbf{y}_{p_n})\right\|_2 + \left\|\mathbf{F}^y\left(\mathbf{y}_n^U\right) - \mathbf{F}^y(\mathbf{y}_{p_n})\right\|_2\right),$$
$$E_n^L = \left\|\mathbf{y}_n^L - \mathbf{y}_{p_n}\right\|_2^2 + \gamma\left(\left\|\mathbf{F}^x\left(\mathbf{y}_n^L\right) - \mathbf{F}^x(\mathbf{y}_{p_n})\right\|_2 + \left\|\mathbf{F}^y\left(\mathbf{y}_n^L\right) - \mathbf{F}^y(\mathbf{y}_{p_n})\right\|_2\right). \quad (8)$$

The above mentioned soft-similarity condition for the candidate patch $\mathbf{y}_{p_n}$ to be included in the dictionary D_n is defined as

$$\mathbf{y}_{p_n} \in D_n \quad \text{if} \quad E_n^d < \kappa E_c^d, \quad (9)$$

where κ sets a threshold on the range of costs for the neighboring patches. Note that such similarity condition plays a similar (rather, a softer) role as neighborhood similarity,[11] which typically advocates that neighbors tend to be similar in appearance. Hence, it is fair to assume that neighboring patches typically tend to have similar respective costs. Essentially, the condition in Eq. (9) helps in excluding those dissimilar neighboring patches, which may not be suitable in group-sparse representation in order to induce a constructive smoothness.

Again, as in the above case, the neighborhood dictionaries in the second stage are also constructed similarly, except with $\mathbf{y}_1^R, \ldots, \mathbf{y}_N^R$ which are the patches neighboring to $\mathbf{y}_c^R$, in the reconstructed slice from the first stage.

$$E_n^d = \left\|\mathbf{y}_n^R - \mathbf{y}_{p_n}\right\|_2^2 + \gamma\left(\left\|\mathbf{F}^x\left(\mathbf{y}_n^R\right) - \mathbf{F}^x(\mathbf{y}_{p_n})\right\|_2 + \left\|\mathbf{F}^y\left(\mathbf{y}_n^R\right) - \mathbf{F}^y(\mathbf{y}_{p_n})\right\|_2\right), \quad (10)$$

where $n = 1 \ldots N$. The soft-similarity condition with respect to the central patch is also considered similar to Eq. (9), except with E_n^d and E_c^d as defined in Eqs. (6) and (10).

In addition, in both the stages, 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 neighboring patches, we do not include the corresponding patch in any dictionary, i.e., the group-structure uses only those patches which satisfy both spatial and appearance based similarity.

3.C. Group-sparse coefficient estimation and patch reconstruction

Having formed the central and neighborhood dictionaries corresponding to a patch to be reconstructed, we solve for sparse coefficients in the $l_{2,1}$ group-sparse representation framework using these dictionaries and corresponding observation patches (from the observed slices). We then use these estimated sparse coefficients to reconstruct the patch in question.

The observation vector $\tilde{\mathbf{y}}_k$ consists of raw intensity as well as gradient features of the observed patch. These are concatenated as $\tilde{\mathbf{y}}_k = [(\mathbf{y}_k^U)^T\ (\gamma\mathbf{F}^x(\mathbf{y}_k^U))^T\ (\gamma\mathbf{F}^y(\mathbf{y}_k^U))^T\ (\mathbf{y}_k^L)^T\ (\gamma\mathbf{F}^x(\mathbf{y}_k^L))^T\ (\gamma\mathbf{F}^y(\mathbf{y}_k^L))^T]^T$ in the first stage, and as $\tilde{\mathbf{y}}_k = [(\mathbf{y}_k^R)^T\ (\gamma\mathbf{F}^x(\mathbf{y}_k^R))^T\ (\gamma\mathbf{F}^y(\mathbf{y}_k^R))^T]^T$ in the second stage. To maintain dimensional consistency with $\tilde{\mathbf{y}}_k$, the atoms in the dictionary are defined as $[(\tilde{\mathbf{y}}_{p_k})^T(\tilde{\mathbf{y}}_{p_k})^T]^T$ in the first stage and $[(\tilde{\mathbf{y}}_{p_k})]$ in the second, where $\tilde{\mathbf{y}}_{p_k} = [(\mathbf{y}_{p_k})^T(\gamma\mathbf{F}^x(\mathbf{y}_{p_k}))^T(\gamma\mathbf{F}^y(\mathbf{y}_{p_k}))^T]^T$.

Instead of using the observation vectors and dictionary directly, we project these onto a PCA-space computed using the dictionary atoms. While PCA provides dimensionality reduction, our primary motivation of using PCA is to achieve further noise robustness. This is because of the observation that the top-most principal components are typically resistance to noise, which is a high-frequency phenomenon and

manifests itself in the lower Eigen-vector space.[24] Such dictionaries projected in the PCA-space, would ideally provide relatively less noisy data to the group-sparse representation problem. This would in turn help in a more faithful coefficient computation. As we demonstrate, while group-sparsity does play a role in mitigating the noise-structure trade-off, using PCA further supports this objective.

Denoting a neighborhood dictionary D_k ($k \in \{\{1, \ldots, N\}, c\}$), of size $M_1 \times M_2$, where M_1 is the dimensionality of feature vectors and M_2 is the number of atoms in the dictionary, we compute a $M_1 \times M_1$ covariance matrix using the mean-subtracted D_k, and subsequently the M_1 eigenvectors. Of these, we choose a small subset of the eigenvectors corresponding to the top-most eigenvalues. Forming a matrix V with the eigenvectors as its columns, we project the dictionary D_k and the observation vector $\tilde{\mathbf{y}}_k$ onto this PCA-space,

$$D_{k_e} = V^T D_{k_\mu},$$
$$\tilde{\mathbf{y}}_{k_e} = V^T \tilde{\mathbf{y}}_{k_\mu}, \tag{11}$$

where D_{k_μ} and $\tilde{\mathbf{y}}_{k_\mu}$ are the mean-subtracted dictionary and observation vector, respectively. D_{k_e} and $\tilde{\mathbf{y}}_{k_e}$ are the corresponding projected dictionary and observation vector, which have a much lesser dimensionality depending on the number of selected eigenvectors. These are chosen so that they contain around 90% of the energy computed using the eigenvalues.

We employ D_{k_e} and $\tilde{\mathbf{y}}_{k_e}$ for our group-sparse representation as

$$\hat{A} = \arg\min_{\alpha_k} \sum_{k \in \{\{1,\ldots,N\},c\}} ||\tilde{\mathbf{y}}_{k_e} - D_{k_e}\alpha_k||_2^2 + \lambda||A||_{2,1}. \tag{12}$$

The above function is optimized via a convex optimization approach proposed in Ref. 21. Once the coefficient matrix $\hat{A}$ is computed, the central patch is reconstructed using the original (preprojected) central dictionary (since the reconstruction needs to be in the original space) as

$$\tilde{\mathbf{y}}_c^O = D_c \hat{\alpha}_c, \tag{13}$$

where $\tilde{\mathbf{y}}_c^O$ is the reconstructed output vector. $\hat{\alpha}_c$ is the estimated coefficient vector (normalized to 1) corresponding to the center patch. The top fraction of $\tilde{\mathbf{y}}_c^O$ (that contains raw intensities) is reshaped and placed into the unknown slice, with overlapping values properly averaged.

Note that although the reconstruction only uses coefficients corresponding to the central patch, the group-sparsity regularization has already acted in estimating these coefficients. The estimated coefficients are constrained so as to select those dictionary patches which agree most with their neighbors and better support controlled smoothness and noise reduction.

3.D. Concerning the two-stage strategy

The need for the two-stage strategy arises from the fact that we focus on better structural correctness rather than localization in the first stage. In this regard, we employ a larger scale of patches, inspired by Zhang *et al.*,[15] which yields structurally more correct results. However, this yields a lack in localization due to smoothing induced by the choice of the larger scale. To minimize further smoothing due to patch averaging during reconstruction, we keep high sparsity in the first stage so that only a few patches (∼10–20) take part in the reconstruction.

The second stage, which operates at a smaller scale, improves the localization over the reconstruction in the first stage. The structural correctness is not hampered since at this stage the reconstructed slice from the first stage serves as the observation. Here, we also relax the sparsity, so that a good number of patches are used in the reconstruction, and exploit group-sparsity in a better way. This indicates that the group-sparsity method works better given a structurally "good" estimate from the first stage.

Note that the approach of Zhang *et al.*[15] also uses a hierarchical method that reduces scale over iterations. However, in that approach, the outputs at the end of iterations are averaged in a weighted fashion, with a large weight assigned to the output at the largest scale (viz. that used in the first stage). In absence of such an averaging, the effect of noise and artifacts increases as the scale is reduced. On the other hand, our second stage operates strictly at a lower scale. This results in better structural localization, while still maintaining robustness to noise and artifacts due to the use of group-sparsity regularization, as we would observe in the experimental results. The pseudo-code of our overall approach is provided in Algorithm 1.

4. EXPERIMENTAL RESULTS

We now provide some results for our proposed method. Our experiments involve the publicly available DIR-Lab lung data,[5] containing ten cases, each with ten respiratory phases. The in-plane resolution for each slice is 1 mm and the superior-inferior resolution is 2.5 mm. We further subsample this data by removing alternate slices, so that we can compare our estimated intermediate slice with the true slice. We provide qualitative as well as quantitative results. We also compare with bicubic interpolation and the method in Ref. 15. The comparison with the latter is due to the observation that our approach can be considered as an extension to that in Ref. 15, as mentioned earlier.

The quantitative metrics that we compute are root-mean-square (RMS) and structural similarity (SSIM).[25] The latter gives an indication of structural improvement and localization, and agrees with qualitative human perception.[25] For sake of clarity, we state that RMS reduces and SSIM increases with the improvement in the results. Also, the SSIM metric ranges between 0 and 1.

The metrics are computed over the lung field mask. An example is as shown in Fig. 4(c), for the lung image, Fig. 4(b). Moreover, as our work focuses on resolution enhancement, we select a range of slices which show good amount of vessel structures. These slices are typically the 30–40 axial slices around the interior of the lung. Thus, after subsampling, for each case (with ten phases), the number of slices range from 130 to 200. Totally, we have over 1700 slices for validation over all ten cases.

ALGORITHM I. Proposed approach for reconstructing intermediate axial slices in 4D-CT.

Given: 4D-CT data $I = \{I_i(s) | i = 1, \ldots, P; s = 1, \ldots, S\}$
Output: slice $I_i^O(s + 0.5)$

Stage 1:
for Loop over voxel locations (x, y) in certain steps
Patches $\mathbf{y}_c^U$ and $\mathbf{y}_c^L$ at (x, y) in slices $I_i(s)$ and $I_i(s + 1)$.

Select overlapping neighboring patches $\mathbf{y}_n^U$ and $\mathbf{y}_n^L$ to the patch at (x, y) in slices $I_i(s)$ and $I_i(s + 1)$.

Construct D_c with candidate patches y_{p_c} across phases selected via Eqs. (4) and (5).

Construct D_n with patches neighboring to selected y_{p_n} if they satisfy Eq. (9), and remove patches from D_c corresponding to those y_{p_n} which do not satisfy the condition.

Compute a PCA on the dictionaries to yield new projected dictionaries D_{k_e} and observation vectors $\tilde{\mathbf{y}}_{k_e}$

Estimate the group-sparse coefficients with via Eq. (12).

Reconstruct the patch $\tilde{\mathbf{y}}_c^O$

Add the intensity values from $\tilde{\mathbf{y}}_c^O$ to reconstructed slice $I_i^R(s + 0.5)$ at (x, y) with proper averaging
end for

Proceed to stage 2 with the reconstructed slice $I_i^R(s + 0.5)$

Stage 2:
for Loop over voxel locations (x, y) in certain steps
Patch $\mathbf{y}_c^R$ at (x, y) in the estimated slice $I_i^R(s + 0.5)$ at a smaller scale.

Repeat steps 2 to 6 as in Stage 1, using patch $\mathbf{y}_c^R$

Add the intensity values from $\tilde{\mathbf{y}}_c^O$ to reconstructed slice $I_i^O(s + 0.5)$ at (x, y) with proper averaging
end for

Also, the quantitative metrics are computed in regions around vessel structures. The structures are extracted using the method in Ref. 26. Moreover, to account for false positives as well as false negatives in computing the quantitative metrics, we extract the vessel regions for both the ground-truth slice and the estimated slice and compute an ORing of the two. We then dilate this ORed map by a small amount to consider some pixels around the exact vessel regions. An example of such a vessel map is shown in Fig. 4(d), computed using the estimated slice and the ground-truth slice in Figs. 4(a) and 4(b), respectively. We believe that such an error computation over vessel regions is rather fitting in a work on super-resolution—an area which primarily focuses on aspects such as localization and noise robustness.

4.A. Parametric variations

We first show the behavior of our approach with respect to some important parameters. Some of the parameters in our method are fixed, inspired by Zhang *et al.*[15] These are $\gamma = 0.2$, $\epsilon = 1.1$, and the patch size, which is 32×32 in the first stage and 16×16 in the second. The value of κ is chosen so that the number of patches in our dictionary is around 400. The 3D region to search for 2D patches from other phases is $11 \times 11 \times 11$, and we employ $N = 4$ neighbors to the center patch in our group-sparsity formulation. For computing the sparse coefficients, we use the MALSAR package.[27] As sparse representation is central to our framework, we test the performance with respect to the sparsity parameter λ. Moreover, as mentioned above, since our rationale of using PCA is motivated toward attaining noise resistance, it is also important to gauge the performance variation with respect to the number of principal components.

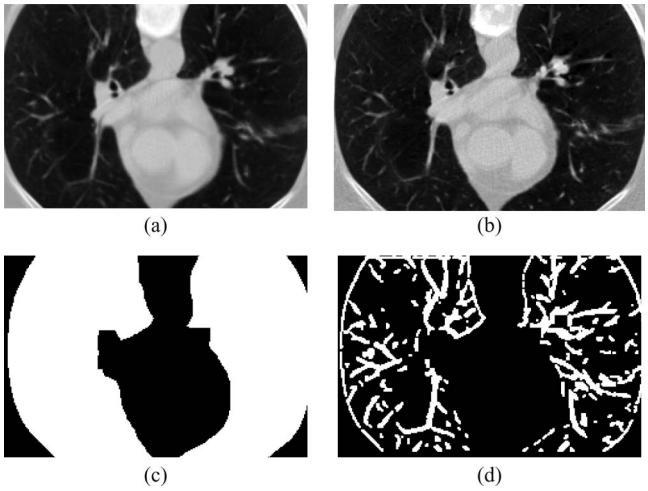
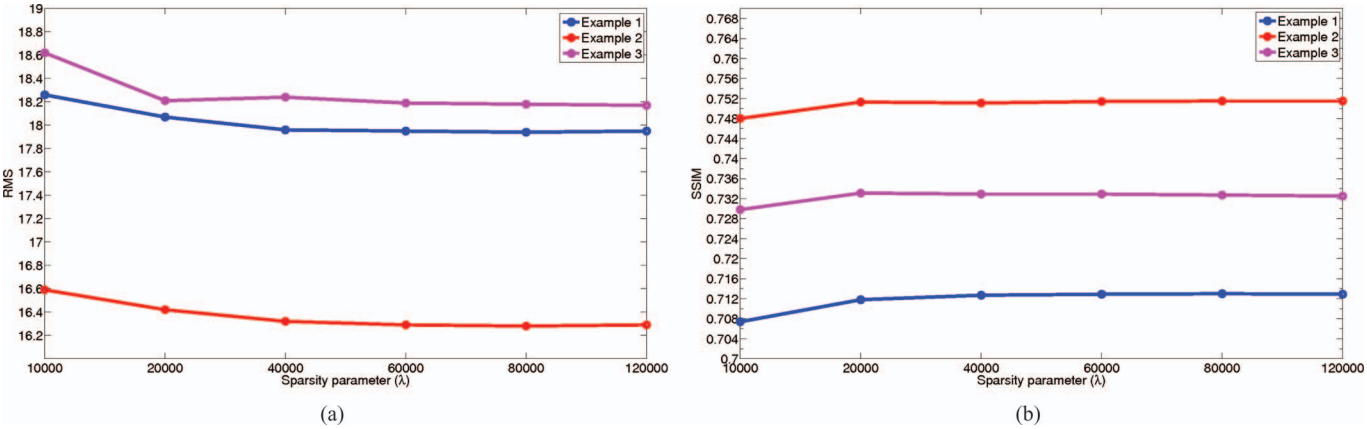


FIG. 4. (a) Estimated slice, (b) ground-truth slice, (c) lung mask region, and (d) vessel regions computed using slices shown in (a) and (b).

Figure 5 shows the variation of average RMS and SSIM over all slices with respect to the sparsity parameter λ, for three examples corresponding to case 1, case 2, and case 5 out of the ten patient cases. The graphs in Fig. 5 is for our approach when computed with the dictionaries *without* the PCA transformation. In this case, $\lambda = 10\,000$ corresponds to a sparsity level of about 60% (i.e., about 60% of coefficients are nonzero). $\lambda = 80\,000$ corresponds to a sparsity level of 20%–30%. The variation indicates that the overall accuracy is somewhat less (in terms of RMS and SSIM), when the number of patches used for reconstruction is large, but improves with the reduction in the number of patches. This is natural since a very large number of patches would also include some outliers and result in oversmoothing. However, it is also observed that after some point (after $\lambda = 80\,000$), the performance remains constant or reduces slightly. Although the sharpness improves, as the number of patches keeps reducing, some useful interphase structural information is also excluded at some point. Thus, considering the variation in both RMS and SSIM in Figs. 5(a) and 5(b), respectively, we infer that the range for optimal performance could be considered around $\lambda = 80\,000$.

In Fig. 6, we show a similar variation with respect to the sparsity parameter, but this time computed with the dictionary in the PCA space (which is our complete approach). Here, we have used the number of eigenvectors to be 200, where the



(a) (b)

FIG. 5. Variation with sparsity parameter λ for the case without PCA. (a) RMS and (b) SSIM.

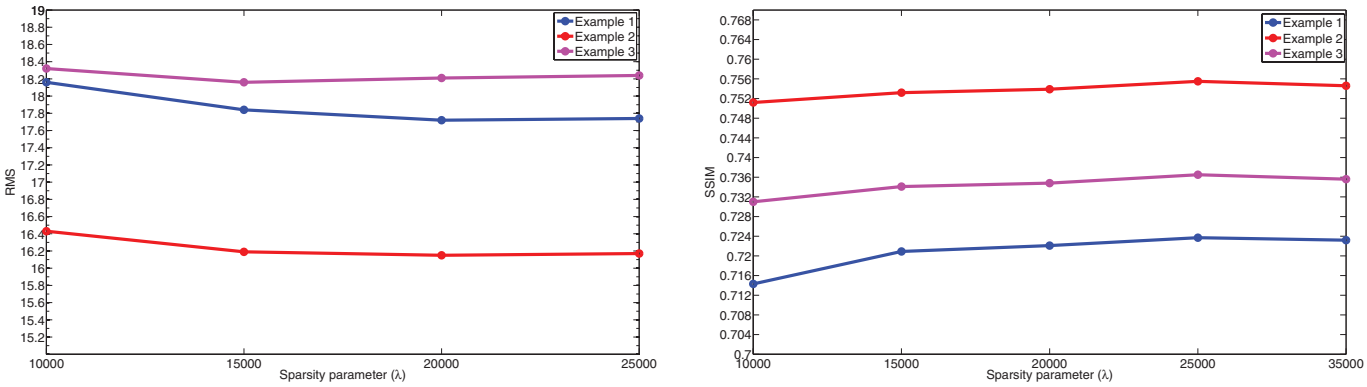
dimensionality of our complete space is over 700 (which is the actual dimensionality of the nonprojected features). Note that in Fig. 6, the value of sparsity parameters is different. This is because we observe that to achieve the similar sparsity in the PCA space, the sparsity parameters have a lower value. Again, we see a similar variation as in Fig. 5, where there is an improvement followed by saturation/reduction in the performance. However, in general, one can observe an improvement in the PCA-based error metrics than in the previous case (RMS is lower and SSIM is higher). This further indicates that the objective of PCA for improvement in noise reduction, is served. Again, based on the RMS and SSIM graphs, in this case, we could say that a good range for λ would be $\lambda = 20\,000$ to $\lambda = 25\,000$.

In Fig. 7, we show the performance variation with respect to the number of principal components at a fixed value of λ. Given our motivation to employ PCA to exploit its noise resistance property, we also acknowledge that using fewer principal components also results in an added risk of compromising useful high frequency components. As the number of principal components increases, the performance can improve due to addition of useful high-frequency components. However, a further increase after some point can also result in noisy components affecting the performance negatively. From the RMS and SSIM variation in Fig. 7, we infer that a good choice for balance between these trade-offs is using around 200 eigenvectors. Thus, we can still employ a far lesser dimensionality than that of the complete space, to achieve a good performance. As we show next our approach also provides considerable improvements over the state-of-the-art.

4.B. Visual qualitative results and comparisons

Having decided on the parameters as discussed above, we first demonstrate some qualitative slice reconstructions in Fig. 8, which shows some example outputs and the corresponding difference maps with the ground-truth slices. The first to fourth columns (in rows 1, 3, 5), respectively, depict the slice reconstructions obtained using bicubic interpolation, the approach in Ref. 15, our output, and ground truth, respectively. The three columns in rows 2, 4, 6 show the difference maps in the same order.

It can be observed that vessel structures in our outputs are far better localized and more accurate than those in the interpolated approach and also show an improvement over the



(a) (b)

FIG. 6. Variation with sparsity parameter with PCA. (a) RMS and (b) SSIM.

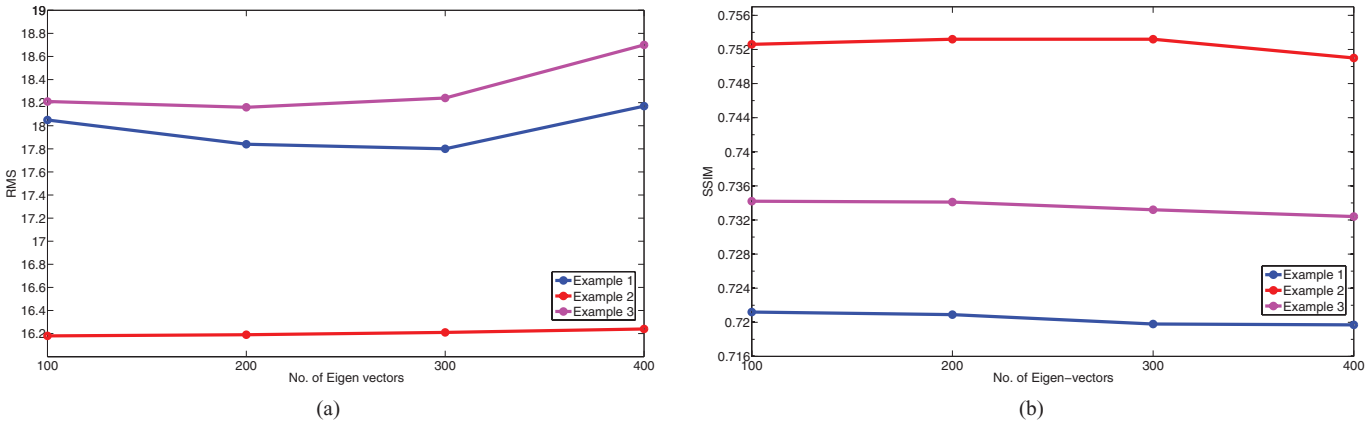


FIG. 7. Variation with the number of eigenvectors. (a) RMS and (b) SSIM.

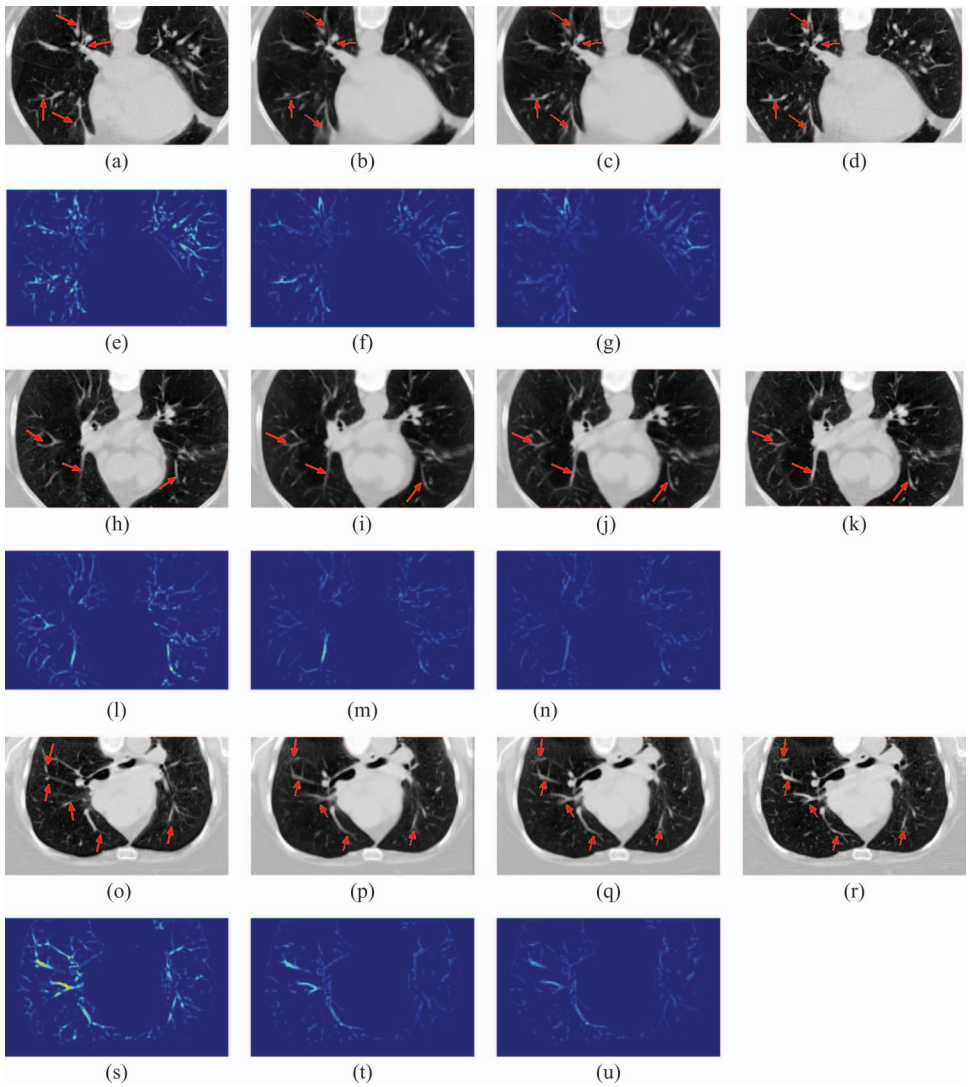


FIG. 8. Examples for reconstruction of some slices: (a, h, o) Outputs using bicubic interpolation with corresponding difference maps shown in (e, l, s). (b, i, p) Outputs using the approach in Ref. 15 with corresponding difference maps shown in (f, m, t). (c, j, q) Outputs using the proposed method with corresponding difference maps shown in (g, n, u). (d, k, r) Ground-truth slices.

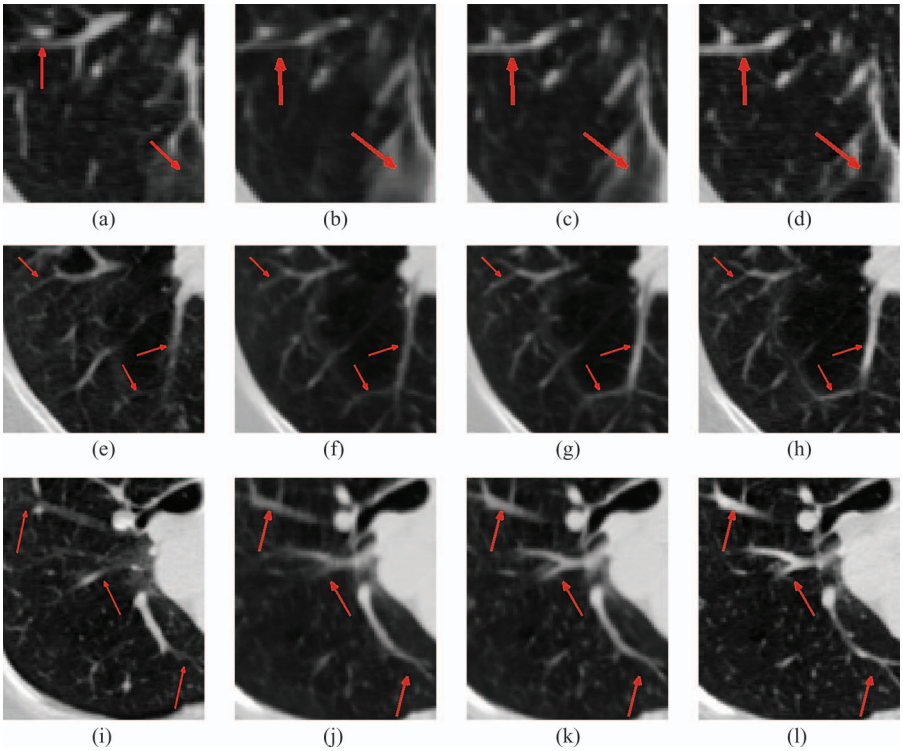


FIG. 9. Close-up views: Each row depicts the close-up views of a region from slices shown in Fig. 8. Columns one to four correspond to the bicubic interpolation, approach in Ref. 15, proposed approach and ground-truth, respectively.

outputs using the method in Ref. 15. Some of the improvements are marked with red arrows. The overall improvements can also be made out via the difference maps, where our approach clearly shows the least differences with the ground-truth among the three.

To depict the improvements better, we zoom into some regions of Fig. 8, and show close-up views in Fig. 9. The rows and columns in Fig. 9 follow the same corresponding order as in Fig. 8. From these close-up views, one can further appreciate the ability of our method in achieving more plausible and better localized structures. As will be seen from the quantitative results, in addition to providing such structural improvements, our approach is also robust to noise and artifacts, yielding an overall good efficacy.

In addition to demonstrating the visual improvements on the above examples, we also highlight the improvements for a clinically more important example, that of involving a tumor. Figure 10 (first, third, and fifth rows) shows examples on three slices from different phases, where the tumor can be seen in the upper-left region (marked by a red arrow). The close-up views focusing on the region around the tumor are shown in second, fourth, and sixth rows. We can see large errors in the shape of the tumor in the bicubic interpolated cases (first column) as compared to the ground-truth (last column). The approach in Ref. 15 (second column) estimates the shape more correctly, but has somewhat blurred boundaries. Our estimation (third column) does a still better job and is nearer to the correct shape in the ground-truth. One can also see some localization improvements in the nearby structures to the tumor (some of which are also shown by arrows). Such improvements in defining the tumor and its nearby structures can help in better planning of the radiation therapy.

4.C. Quantitative results

Finally, we provide quantitative results averaged over complete sets of data used in our experiments. Tables I and II, respectively, provides casewise RMS errors and SSIM for bicubic interpolation, the results reported in Ref. 15, and with our approach for the optimum choice of parameters discussed earlier. Observe that our approach shows a clear improvement with respect to both metrics for all ten cases (note that such an order of improvement is common in contemporary super-resolution.[10,28]) The RMS results support our claims for overall mitigation of the noise-structure trade-off, whereas the SSIM results emphasize better structure enhancement. Also note that in most cases using PCA also shows further improvement. Indeed, for SSIM, the improvement can be seen for all cases.

In Fig. 11 we show the percentage of slices across cases, for which our approach favorably compares with that in Ref. 15 in terms of RMS and SSIM. 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.

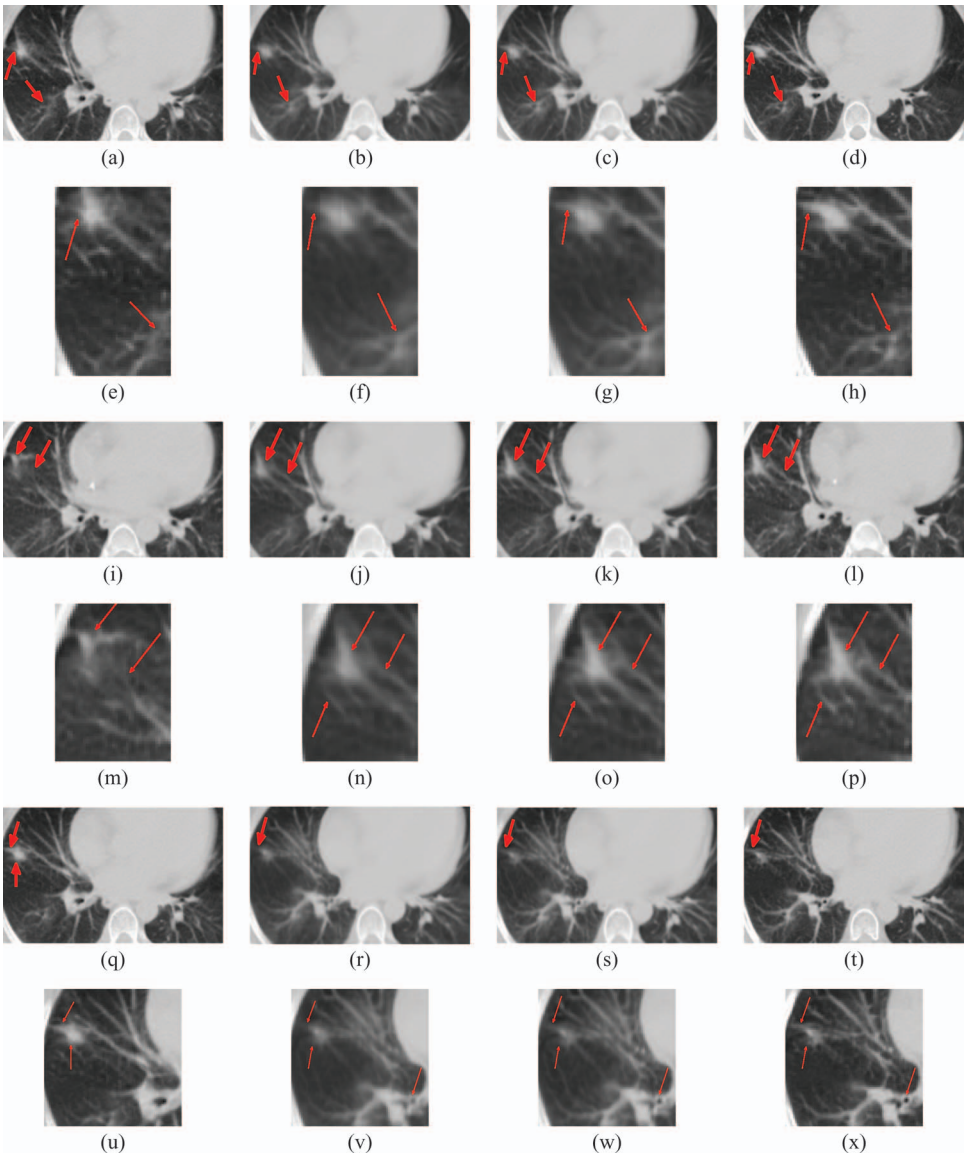


FIG. 10. Three examples of axial slice reconstruction involving tumor: (a, i, q) Outputs using bi-cubic interpolation. (b, j, r) Outputs using the approach in Ref. 15. (c, k, s) Outputs using the proposed method. (d, l, t) Ground-truth slices. Respective zoomed in regions shown in (e, m, u), (f, n, v), (g, o, w), (h, p, x).

TABLE I. Average RMS for ten cases.

Case (No. of slices)	RMS: Bicubic	RMS: (Ref. 15)	RMS: Proposed (w/o PCA)	RMS: Proposed (PCA)
Case 1 (170)	28.97	18.56	17.94	**17.74**
Case 2 (200)	27.83	16.74	16.28	**16.17**
Case 3 (160)	26.82	15.86	15.11	**15.07**
Case 4 (170)	26.95	17.10	16.33	**16.23**
Case 5 (170)	29.00	19.23	**18.18**	18.24
Case 6 (190)	24.81	17.75	17.25	**17.22**
Case 7 (180)	28.12	18.94	**17.90**	17.92
Case 8 (180)	37.38	26.72	24.57	**24.53**
Case 9 (130)	25.34	16.63	**16.01**	16.10
Case 10 (190)	37.10	25.40	23.68	**23.61**
Average (total: 1740)	29.23	19.29	18.33	**18.28**

TABLE II. Average SSIM for ten cases.

Case (No. of slices)	SSIM: Bicubic	SSIM: (Ref. 15)	SSIM: Proposed (w/o PCA)	SSIM: Proposed (PCA)
Case 1 (170)	0.5492	0.7069	0.7130	**0.7237**
Case 2 (200)	0.5777	0.7445	0.7515	**0.7555**
Case 3 (160)	0.5895	0.7589	0.7740	**0.7762**
Case 4 (170)	0.5869	0.7372	0.7552	**0.7587**
Case 5 (170)	0.5671	0.7184	0.7327	**0.7365**
Case 6 (190)	0.5151	0.6255	0.6425	**0.6428**
Case 7 (180)	0.5181	0.6622	0.6862	**0.6875**
Case 8 (180)	0.4913	0.6137	0.6471	**0.6518**
Case 9 (130)	0.5558	0.6980	0.7124	**0.7136**
Case 10 (190)	0.4820	0.6148	0.6469	**0.6518**
Average (total: 1740)	0.5433	0.6860	0.7062	**0.7098**

5. DISCUSSION AND CONCLUSION

In this work, we proposed a novel approach for enhancing the superior-inferior resolution of 4D-CT, by reconstructing missing intermediate slices in a certain phase by exploiting local complimentary sampling information available in the observed slices of other phases, due to lung-motion.

Our work built upon a recent sparse-representation-based work in Ref. 15. We looked to mitigate the trade-off in resolution enhancement between achieving good localization/sharpness of anatomical features as well as good accuracy which is affected by noise/artifacts. Toward this end, we proposed a novel group-sparsity based method to incorporate constraints from similar overlapping patches, which are neighboring to the patch to be reconstructed.

We discussed the regularizing role of group-sparsity which provides constructive smoothness constraints. This allowed a smaller scale of patches (than that in the state-of-the-art[15]) in our patch reconstruction process, and helped in improving the localization without compromising on accuracy. We demonstrated that such an application of group-sparsity allows for better structure preservation and noise resistance. Moreover, we further strengthened noise robustness using the PCA-based transformation of features, which, in principle, embeds the less noisy data variation in the top few principal components, which are used to compute the features in a reduced dimensional space.

We demonstrated, both qualitatively and quantitatively, that the above ideas when formulated in our proposed approach indeed serve our objective of achieving a good localization-accuracy balance. We further conducted an empirical parametric analysis to compute our best results. However, one can also observe that our approach is not too sensitive with respect to the parameter variation. The curves in Figs. 5–7 have a small slope, which means that the performance of our approach is relatively stable over a range of parameter variation. Such a stability with respect to parameters is an important quality for an estimation method.

As mentioned earlier, we believe that this work plays an important role in advancing the relatively young area 4D-CT resolution enhancement. The importance of our work from a clinical perspective can be appreciated, as the improvement in localization and accuracy of anatomical structures in 4D-CT can facilitate better delineation or segmentation. In fact, we also demonstrated the efficacy of our approach on tumor data, where the tumor and its neighboring structures are much better defined using our method, which is a clear indication of its positive implications in radiation therapy.

Also, our work is one of the few to highlight the regularizing role of group-sparsity for image reconstruction. While, group-sparse representation has been popular in the classification and machine learning arena,[29] its exploration for image estimation has been limited. Hence, from the technical point of view, our work could also encourage further research of group-sparsity and other related frameworks for regularization and image estimation.

Thus, our work serves a dual purpose of further contributing to the area of resolution enhancement of lung 4D-CT, and of demonstrating the potential of group-sparse representation for image reconstruction.

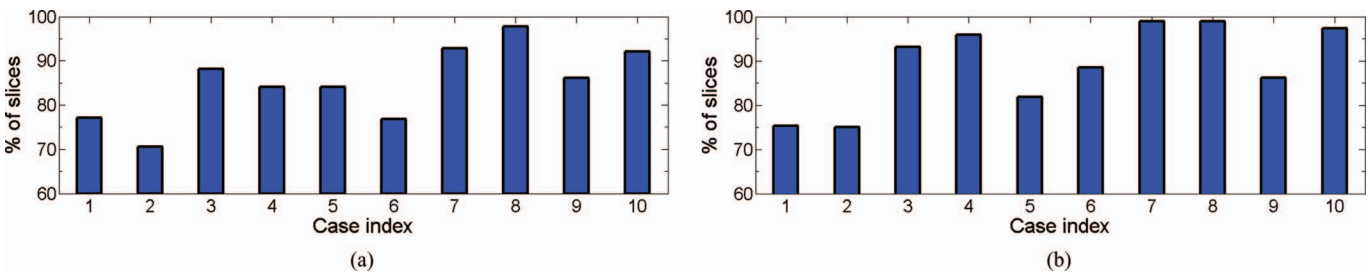


FIG. 11. Percentage of slices on which our approach performs favorably over (Ref. 15) for (a) RMS (b) SSIM.

In future, it would be interesting to explore the role of dictionary learning, and also other multi-task learning frameworks for the task of 4D-CT resolution enhancement. We will also explore the utilization of image registration technique, as we have used in various applications,[30–36] to help estimate the corresponding patches for resolution enhancement.

[a]Electronic addresses: arnav.bhavsar@gmail.com; guorong_wu@med.unc.edu; and dgshen@med.unc.edu
[b]Electronic mail: Jun_Lian@med.unc.edu

[1]R. Li, J. Lewis, L. Cervinno, and S. Jiang, "4D-CT sorting based on patient internal anatomy," Phys. Med. Biol. **54**, 4821–4833 (2009).
[2]J. Ehrhardt, R. Werner, T. Frenzel, D. Saring, W. Lu, D. Low, and H. Handels, "Reconstruction of 4D-CT data sets acquired during free breathing for the analysis of respiratory motion," Proc. SPIE **6144** (2006).
[3]D. Gomez and J. Chang, "Adaptive radiation for lung cancer," J. Oncol. **2011** (2011).
[4]J. Lian and L. Marks, "Adaptive radiation therapy for lung cancer," *Adaptive Radiation Therapy*, edited by X. Allen Li (CRC, Florida, 2011).
[5]R. Castillo, E. Castillo, R. Guerra, V. Johnson, T. McPhail, A. Garg, and T. Guerrero, "A framework for evaluation of deformable image registration spatial accuracy using large landmark point sets," Phys. Med. Biol. **54**, 1849–1870 (2009).
[6]S. Rit, D. Sarrut, and L. Desbat, "Comparison of analytic and algebraic methods for motion-compensated cone-beam CT reconstruction of the thorax," IEEE Trans. Med. Imaging **28**(10), 1513–1525 (2009).
[7]J. Tian and K. Ma, "A survey on super-resolution imaging," Signal Image Video Process. **5**(3), 329–342 (2011).
[8]N. Bose and N. Ahuja, "Superresolution and noise filtering using moving least squares," IEEE Trans. Image Process. **15**(8), 2239–2248 (2006).
[9]S. Farsiu, M. D. Robinson, M. Elad, and P. Milanfar, "Fast and robust multiframe super resolution," IEEE Trans. Image Process. **13**(10), 1327–1344 (2004).
[10]J. Yang, J. Wright, T. S. Huang, and Y. Ma, "Image super-resolution via sparse representation," IEEE Trans. Image Process. **19**(11), 2861–2873 (2010).
[11]W. T. Freeman, T. R. Jones, and E. C. Pasztor, "Example-based super-resolution," IEEE Comput. Graphics Appl. **22**(2), 56–65 (2002).
[12]R. Z. Shilling, T. Q. Robbie, T. Bailloeul, K. Mewes, R. M. Mersereau, and M. E. Brummer, "A super-resolution framework for 3-D high-resolution and high-contrast imaging using 2-D multislice MRI," IEEE Trans. Med. Imaging **28**(5), 633–644 (2009).
[13]A. Rueda, N. Malpica, and E. Romero, "Single-image super-resolution of brain MR images using overcomplete dictionaries," Med. Image Anal. **17**, 113–132 (2013).
[14]G. Wu, J. Lian, and D. Shen, "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).
[15]Y. Zhang, G. Wu, P. Yap, Q. Feng, J. Lian, W. Chen, and D. Shen, "Hierarchical patch-based sparse representation - A new approach for resolution enhancement of 4D-CT lung data," IEEE Trans. Med. Imaging **31**(11), 1993–2005 (2012).
[16]N. Nguyen, P. Milanfar, and G. Golub, "A Computationally efficient super-resolution image reconstruction algorithm," IEEE Trans. Image Process. **10**(4), 573–583 (2001).
[17]S. Babacan, R. Molina, and A. Katsaggelos, "Total variation super resolution using a variational approach," Int. Conf. Image Process. (ICIP) 641–644 (2008).
[18]J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman, "Non-local sparse models for image restoration," Int. Conf. Comput. (ICCV) 2272–2279 (2009).
[19]C. Yang, J. Huang, and M. Yang, "Exploiting self-similarities for single frame super-resolution," Asian Conf. Comput. (ACCV) 497–510 (2010).
[20]Y. Zhang, G. Wu, P. Yap, Q. Feng, J. Lian, W. Chen, and D. Shen, "Non-local means resolution enhancement of lung 4D-CT data," Med. Image Comput. Comput. Assist. Interv. (MICCAI 2012) **7510**, 214–222 (2012).
[21]J. Liu, S. Ji, and J. Ye, "Multi-task feature learning via efficient $l_{2,1}$-norm minimization," Conference on Uncertainty in Artificial Intelligence (UAI) 339–348 (2009).
[22]G. Obozinski, B. Taskar, and M. Jordan, "Multi-task feature selection," *Workshop of Structural Knowledge Transfer for Machine Learning in the 23rd International Conference on Machine Learning (ICML)* (2006).
[23]Note that a special case of Eq. (2) for a single-task would be same as the traditional sparse-representation problem of (1). A more formal discussion on $l_{2,1}$-group-sparsity can be found in Refs. 21 and 22.
[24]A. K. Jain, *Fundamentals of Digital Image Processing* (Prentice Hall, India, 1989).
[25]Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: From error visibility to structural similarity," IEEE Trans. on Image Processing **13**(4), 600–612 (2004).
[26]J. Cheng, C. Chen, E. Cole, E. Pisano, and E. D. Shen, "Automated delineation of calcified vessels in mammography by tracking with uncertainty and graphical linking techniques," IEEE Trans. Med. Imaging **31**(11), 2143–2155 (2012).
[27]J. Zhou, J. Chen, and J. Ye, "MALSAR: Multi-tAsk Learning via StructurAl Regularization," (Arizona State University 2012).
[28]H. Takeda, P. Milanfar, M. Protter, and M. Elad, "Super-resolution without explicit subpixel motion estimation," IEEE Trans. Image Process. **18**(9), 1958–1975 (2009).
[29]X. Yuan, X. Liu, and S. Yan, "Visual classification with multitask joint sparse representation," IEEE Trans. Image Process. **21**(10), 4349–4360 (2012).
[30]D. Shen and C. Davatzikos, "HAMMER: hierarchical attribute matching mechanism for elastic registration," IEEE Transactions on Medical Imaging **21**(11), 1421–1439 (2002).
[31]Y. Zhan, M. Feldman, J. Tomaszeweski, C. Davatzikos, and D. Shen, "Registering histological and MR images of prostate for image-based cancer detection," Medical Image Computing and Computer-Assisted Intervention–MICCAI 2006, 620–628 (2006).
[32]D. Shen, W. Wong, and H. Horace, "Affine-invariant image retrieval by correspondence matching of shapes," Image and Vision Computing **17**(7), 489–499 (1999).
[33]G. Wu, F. Qi, and D. Shen, "Learning-based deformable registration of MR brain images," IEEE Transactions on Medical Imaging **25**(9), 1145–1157 (2006).
[34]D. Shen, Z. Lao, J. Zeng, W. Zhang, I. Sesterhenn, L. Sun, J. Moul, E. Herskovits, G. Fichtinger, and C. Davatzikos, "Optimized prostate biopsy via a statistical atlas of cancer spatial distribution," Medical Image Analysis **8**(2), 139–150 (2004).
[35]H. Qiao, H. Zhang, Y. Zheng, D. Ponde, D. Shen, F. Gao, A. Bakken, A. Schmitz, H. Kung, V. Ferrari, and R. Zhou, "Embryonic stem cell grafting in normal and infarcted myocardium: Serial assessment with MR imaging and PET dual detection," Radiology **250**(3), 821–829 (2009).
[36]E. Zacharaki, D. Shen, S. Lee, and C. Davatzikos, "ORBIT: A multiresolution framework for deformable registration of brain tumor images," IEEE Transactions on Medical Imaging **27**(8), 1003–1017 (2008).