A method for volumetric imaging in radiotherapy using single x-ray projection

Yuan Xu

*Department of Radiation Oncology, University of Texas Southwestern Medical Center, Dallas, Texas 75235 and
Department of Biomedical Engineering, Southern Medical University, Guangzhou, Guangdong 510515, China*

Hao Yan, Luo Ouyang, and Jing Wang

Department of Radiation Oncology, University of Texas Southwestern Medical Center, Dallas, Texas 75235

Linghong Zhou

Department of Biomedical Engineering, Southern Medical University, Guangzhou, Guangdong 510515, China

Laura Cervino

*Center for Advanced Radiotherapy Technologies and Department of Radiation Medicine and Applied Sciences,
University of California San Diego, La Jolla, California 92037*

Steve B. Jiang^{a)} and Xun Jia^{a)}

Department of Radiation Oncology, University of Texas Southwestern Medical Center, Dallas, Texas 75235

(Received 23 October 2014; revised 3 March 2015; accepted for publication 7 April 2015;
published 21 April 2015)

Purpose: It is an intriguing problem to generate an instantaneous volumetric image based on the corresponding x-ray projection. The purpose of this study is to develop a new method to achieve this goal via a sparse learning approach.

Methods: To extract motion information hidden in projection images, the authors partitioned a projection image into small rectangular patches. The authors utilized a sparse learning method to automatically select patches that have a high correlation with principal component analysis (PCA) coefficients of a lung motion model. A model that maps the patch intensity to the PCA coefficients was built along with the patch selection process. Based on this model, a measured projection can be used to predict the PCA coefficients, which are then further used to generate a motion vector field and hence a volumetric image. The authors have also proposed an intensity baseline correction method based on the partitioned projection, in which the first and the second moments of pixel intensities at a patch in a simulated projection image are matched with those in a measured one via a linear transformation. The proposed method has been validated in both simulated data and real phantom data.

Results: The algorithm is able to identify patches that contain relevant motion information such as the diaphragm region. It is found that an intensity baseline correction step is important to remove the systematic error in the motion prediction. For the simulation case, the sparse learning model reduced the prediction error for the first PCA coefficient to 5%, compared to the 10% error when sparse learning was not used, and the 95th percentile error for the predicted motion vector was reduced from 2.40 to 0.92 mm. In the phantom case with a regular tumor motion, the predicted tumor trajectory was successfully reconstructed with a 0.82 mm error for tumor center localization compared to a 1.66 mm error without using the sparse learning method. When the tumor motion was driven by a real patient breathing signal with irregular periods and amplitudes, the average tumor center error was 0.6 mm. The algorithm robustness with respect to sparsity level, patch size, and presence or absence of diaphragm, as well as computation time, has also been studied.

Conclusions: The authors have developed a new method that automatically identifies motion information from an x-ray projection, based on which a volumetric image is generated. © 2015 American Association of Physicists in Medicine. [<http://dx.doi.org/10.1118/1.4918577>]

Key words: tumor motion management, time-resolved volumetric image, lung motion model, sparse learning, intensity baseline correction

1. INTRODUCTION

Time-resolved volumetric imaging is a key component for motion management in lung radiotherapy.^{1,2} A number of solutions have been proposed to provide high-quality images at different stages of radiotherapy treatment, such

as 4D-CT for planning,^{3–6} 4D or motion-compensated cone beam CT (CBCT) for pretreatment setup,^{7–19} and various online imaging methods during treatment.^{20–29} In particular, real-time imaging, i.e., instant volumetric patient image reconstruction, has recently attracted significant research attention for the purpose of visualizing patient anatomy during

treatment for many clinical tasks such as tumor tracking, treatment monitoring, and delivered dose assessment.

A volumetric image may be derived from a variety of different measurements. It is of particular research interest to study single x-ray projection-based volumetric image reconstruction because of the wide availability of CBCT systems integrated in medical linear accelerators.^{30–32} However, it is not easy to achieve this goal. At a given moment, one can only measure one 2D projection of the 3D volumetric image, and the voxel values along the x-ray projection direction are integrated. Additional information must be supplied to retrieve each voxel value. One such type of prior information is a patient-specific lung motion model.^{22,26–28,33} We have previously proposed a principal component analysis (PCA)-based lung motion model. In this model, the principal components of the deformation vector fields (DVF) for lung motion are identified and the first few of them are found to be sufficient to represent the entire lung motion, which greatly reduces the degrees of freedom of the problem.^{22,24,25,34} This concise representation of lung motion aids the development of a real-time volumetric imaging technique, as the limited amount of measurements at a given moment are now sufficient to determine the PCA coefficients. Based on this idea, we previously proposed a method, and for the first time demonstrated the feasibility of reconstructing instantaneous volumetric images based on a single x-ray projection measurement.²⁴ Specifically, the PCA coefficients of the lung model were adjusted through an optimization process so that the forward calculated projection of the volumetric image, corresponding to the PCA coefficients, matched the measurements. A similar concept has also been employed in 4D-CBCT reconstruction where multiple projections corresponding to a given breathing phase are used.²⁷

Despite these successes, there remain two major issues associated with this optimization-based approach for real-time imaging. First, the forward matching process tries to minimize the intensity difference between the calculated and the measured projections. It is not robust with respect to intensity mismatch between the two due to the neglected x-ray scatter in the forward projection calculations. The method works for some patients, but its robustness in general needs to be improved. Second, for each measured x-ray projection, a non-linear optimization problem is solved via a certain iterative algorithm. It is very computationally challenging to achieve high performance so that the volumetric image can be updated at a sufficiently high frequency rate to be considered real-time.

In this paper, we propose a new method to achieve time-resolved volumetric imaging that solves the two aforementioned problems. As opposed to an optimization-based approach, our method realizes the goal via a mapping-based approach. A relationship between the x-ray projections and the corresponding volume is first established via a learning process. The learned mapping is applied to a measured projection, yielding the volumetric image. Because only a simple forward calculation is needed to generate the image based on the projection, computation efficiency is ensured. We have also developed intensity correction methods and incorporated sparse learning techniques to achieve system robustness.

2. METHODS AND MATERIALS

Generally speaking, there exists a forward mapping between a 3D patient anatomy and an x-ray projection $p = P[f]$, where p is the projection image and f is the 3D volumetric data. Note this relation has been simplified to a matrix multiplication form $p = Pf$ in the context of CBCT reconstruction, where P is a projection matrix. However, in reality, the mapping is highly complicated due to numerous physical processes involved in the data acquisition process, e.g., scatter, beam hardening, detector response, etc. Hence, it is quite difficult to invert the mapping, i.e., retrieve the volumetric image f based on projection images. Especially when there is only one projection image available, it is impossible to reconstruct the image utilizing conventional CBCT reconstruction methods.

The fundamental concept behind our approach is to employ a sparse learning technique to generate an inverse mapping in which $f = P^{-1}[p] \equiv M[p]$. Specifically, our model is characterized by the following. (1) We utilize a deformation vector domain, where $f(x) = f_0(x + v)$. Here, v is a vector field that deforms the prior image f_0 . This approach allows us to utilize the PCA-based lung motion model,^{22,25,34} under which v is linearly related to the PCA coefficient w . This substantially reduces the degrees of freedom, as it has been demonstrated that only a few PCA components are sufficient to describe the lung motion to a satisfactory level of accuracy. (2) We then establish a mapping between the set of PCA coefficients w and a projection image p , $w = X[p]$. In general, the function $X[\cdot]$ is expected to be complex and nonlinear. However, a linear function may be used to approximate this function and to capture the major relationship between the two variables. Specifically, we take $w = Xp$, where X is simply a matrix to be determined through a sparse learning process using prior 4D-CT data of the patient. The validity of this assumption will be to a certain extent demonstrated by the successful experimental studies.

Based on this idea, we propose a method with a workflow shown in Fig. 1. Our method consists of two stages: model training and model application. In the training stage, we first obtain motion PCA coefficients by analyzing the DVFs obtained through deformable image registration (DIR) of the 4D-CT images. Meanwhile, we compute x-ray projections

FIG. 1. Workflow of the proposed algorithm.

of the 4D-CT. The corresponding PCA coefficients and projections are fed to a sparse learning algorithm, yielding a motion model. At the application stage, once a measured projection is available, it is first processed to remove intensity inconsistency between the measured projections and the simulated ones. After that, the trained model is applied, yielding the PCA coefficients corresponding to the measured projection. A volumetric image is then derived according to these PCA coefficients.

In the rest of this section, we will first briefly review the PCA representation of the lung motion model. We will then present our method to build the mapping. To demonstrate the feasibility of this approach, we focus our study on a scenario in which all projections are taken at a fixed projection angle. This approach could potentially be applied to imaging the lung area during the delivery of intensity modulated radiation therapy (IMRT), where the linear accelerator gantry is fixed at one angle during beam delivery, as are the x-ray source and the detector.

2.A. PCA-based lung motion model

Suppose there are N phases of 4D-CT images in one breath cycle. The DVFs between the N phases and a reference phase (e.g., the end of the exhale phase) can be obtained by deformable image registration such as demons algorithm.³⁵ The resulting DVFs for these phases form a matrix V , where each column represents the whole DVF field for a corresponding phase. Conducting PCA on this matrix yields

$$v_n \approx \bar{v} + \sum_{k=1}^K u_k w_{k,n}, \quad n = 1, 2, \dots, N, \quad (1)$$

where u and w represent PCA eigenvectors and coefficients, respectively. The subscript k is the index for the principal component, and K is the number of principal components that one selects to approximate the motion vector field. Eliminating components with large k values also removes noises and inaccuracy in the vector field that may be caused by registration error. According to previous studies,^{22,25} $K = 2-3$ is adequate to accurately describe respiratory motion. We have therefore used $K = 3$ throughout our work. In a matrix form, Eq. (1) can be written as

$$v = [A \quad \bar{v}] \begin{bmatrix} w \\ 1 \end{bmatrix}, \quad (2)$$

where A is a matrix consisting of principal components.

2.B. Projection partitioning

To generate the projection data used to build the model, we computed an X-ray projection for each 4D-CT image at the given beam angle using Siddon's ray-tracing algorithm.³⁶ We wanted to maximally extract information from these projections and correlate it with the motion vector field. As such, we chose to use patches on the projection image to predict motion, as opposed to using the entire image. The rationale behind this strategy comes from the observation

that the projection contains both areas with large motion, such as those around diaphragm and chest wall, and areas with small or no motion, such as air outside the patient body. The latter offers no extra motion information and may even "dilute" the motion information provided by those areas with strong motion information, reducing prediction accuracy and robustness. Hence, we elected to use only a few small rectangular areas for motion prediction purposes. As such, we partitioned the projection image with two sets of interlaced rectangular grids as illustrated in Fig. 2. Assuming the projection image resolution to be $U \times V$ pixels, we partitioned it into small patches, each of size $u \times v$ pixels. In order to capture the motion information, we partitioned the projection twice with two grids, one starting at the location $[1, 1]$ and the other starting at $[u/2, v/2]$. Note that there are many other ways to partition the projection image. However, for simplicity we used only the two grids mentioned here, and these two interlaced grids were found to be sufficient in our studies.

2.C. Sparse learning model

As mentioned previously, different patches have different degrees of correlation with the PCA coefficients of the DVFs. There are two issues one needs to consider when building the correlation model in this training stage: (1) selecting those patches that are useful in terms of predicting motion, and (2) building the motion model. We opted to address these two issues simultaneously via a sparse learning approach.

For the projection image at a given phase n , we extracted a long column vector d^n of length $J \equiv H \times M$, where $H = u \times v$ is the number of pixels per patch, and M is the number of patches considered. The vector d^n consists of M subvectors. Each of them is of length H and corresponds to the H intensity values in a patch. These column vectors d^n form a matrix $D := [d^1; d^2; \dots; d^N] \in \mathbb{R}^{J \times N}$. Meanwhile, we also used PCA coefficients $w_{k,n}$ of the DVFs to form a matrix $W \in \mathbb{R}^{K \times N}$, where each column of W is for one breathing phase n . Our assumption of a linear relationship between the projection image and the PCA coefficients translated to the linear equation $XD = W$, with $X \in \mathbb{R}^{K \times J}$. Moreover, the matrix X can be partitioned into M sub-matrices $[X_1; X_2; \dots; X_M]$, and $X_m \in \mathbb{R}^{K \times H}$ represents the patch m . If there is a nonzero entry in the submatrix X_m , it implies that the patch corresponding to this submatrix contributes to the motion model.

Under this structure, the assumption that only a few patches are important to predict motion indicates that only a few submatrices of X will be nonzero. Hence, our next purpose was to find an X matrix with this desired structure, such that $XD = W$. As such, we defined an objective function $\sum_{m=1}^M \|X_m\|_F$, where $\|\cdot\|_F$ denotes the matrix Frobenius norm. It has been demonstrated previously that minimizing the objective function of this form encourages the sparseness of the solution matrix among different submatrices, whereas the sparse requirement within each submatrix is not imposed.³⁷ The application of a norm defined as such has been recently explored in many problems such as beam orientation optimization for IMRT (Ref. 38) and surface marker locations selection for tumor motion estimation.³⁹

FIG. 2. Illustration of how a projection image is partitioned into small rectangular patches by two grids shown in (a) and (b).

With this objective function defined, we formulate our problem as

$$X^* = \arg \min_X \sum_{m=1}^M \|X_m\|_2, \quad \text{s.t. } XD = W. \quad (3)$$

In reality, since the training data always contain noise, we solve an unconstrained optimization problem instead

$$X^* = \arg \min_X \frac{1}{2} \|XD - W\|_2^2 + \lambda \sum_{m=1}^M \|X_m\|_2, \quad (4)$$

to allow deviations of the predicted PCA coefficients X^*D from the training data W . The positive constant λ is manually selected to balance the fidelity term $1/2 \|XD - W\|_2^2$ and the sparse term $\sum_{m=1}^M \|X_m\|_2$. To solve this optimization problem, we use the “glLeastR” algorithm in the package developed by Liu *et al.*⁴⁰

2.D. Measured projection preprocess

Before using the trained model to predict a volumetric image based on a single projection image, there is one more problem to solve. The projection images in the aforementioned training stage are generated by a ray-tracing algorithm and are free of scatter. Yet the acquired projection, based on which a volumetric image is to be determined, comes from measurements directly. There is an intensity inconsistency between the measured projections and those in the training stage. A baseline correction step has to be performed to correct the intensity of the measured projections.

For this purpose, we developed the following strategy. We first acquired a number of projections over one breath cycle and averaged these projection images. We then partitioned the averaged projections into patches in the same way as presented in Sec. 2.B. Meanwhile, we averaged the projection images used in the training stage. These two average images differed by a background. We chose to estimate a linear transformation for each patch such that, when applied to the patch in the measured projection, the resulting patch and the corresponding patch in the training stage had the same mean value and standard deviation. Ideally, we would like to match the histograms of the two patches. But practically it was found that matching only the first two statistical moments, i.e., mean value and standard deviation, yielded satisfactory background

correction. Specifically, for a patch m with q_m denoting the vector containing its intensity values, the transformed patch becomes p_m ,

$$p_m = a(m)q_m + b(m), \quad (5)$$

where $a(m)$ and $b(m)$ are the parameters for this patch to be estimated. Using the patch intensity vector from the averaged training projection $\bar{p}^m$ and that from the measured projections $\bar{q}^m$, it can be derived⁴¹ that

$$a(m) = \text{STD}(\bar{p}^m) / \text{STD}(\bar{q}^m) \quad (6)$$

$$b(m) = E(\bar{p}^m) - a(m)E(\bar{q}^m). \quad (7)$$

Here, $E(\cdot)$ and $\text{STD}(\cdot)$ denote the calculations of mean value and standard deviation of a vector, respectively.

2.E. Generating a volumetric image with one measured projection

With the model trained and the patch intensity baseline correction parameters, we are ready to generate a volumetric image based on one measured projection image. Specifically, once a projection image is acquired, the patches that are relevant to the motion prediction are selected. Depending on their locations, corresponding intensity correction parameters are used to adjust the patch intensities. The resulting patches are then multiplied with the trained X^* matrix, yielding the corresponding PCA coefficients of the deformation vector field. A vector field is generated, which is finally applied to the reference image, yielding the final volumetric image corresponding to the projection.

2.F. Evaluations

2.F.1. Synthetic study

To evaluate our algorithm, we generated a test dataset based on 4D-CT data of a real patient case. Specifically, the 4D-CT images were used in the training stage as presented previously. To test our algorithm, volumetric images at a much higher temporal resolution were needed. As such, we first computed ten DVFs between a reference phase (end of exhale) and each of the ten phases of the 4D-CT using a demons deformable image registration algorithm. Then we interpolated six new DVFs between two adjacent phases by cubic spline interpola-

FIG. 3. Patch selection result. In (a) and (b), the patches indicated by the rectangles are selected by sparse learning algorithm.

tion. A total of 60 DVFs were applied to the reference image, yielding 60 volumetric images corresponding to the variations of patient anatomy during a breathing cycle.

Corresponding to those 60 images in a breathing cycle, we synthesized x-ray projections using an in-house developed software tool (gDRR)⁴² under the geometry of a Varian TrueBeam On-Board Imaging system (Varian Medical System, Inc., Palo Alto, CA). In the gDRR package, the projections are calculated using a Monte Carlo method with various features in a realistic scan considered, such as energy spectrum, source fluence map, and detector response, as well as scatter and quantum noise. The x-ray images have a resolution of 512×384 pixels with a physical size of 40×30 cm². These simulated projections were then used to generate corresponding volumetric images using our algorithm.

For this synthetic data set, we have the ground truth of the PCA coefficients of the DVFs, as well as the images. These will be used to evaluate the results generated by our algorithm.

2.F.2. Experimental study

We also tested our algorithm on a moving lung phantom. The phantom is constructed with an air cavity and materials equivalent to natural bone and soft tissue. A cylindrical tumor

~ 3 cm in diameter and ~ 3 cm in height was attached to the end of a Styrofoam rod, which was put inside the chest cavity and programmed to move in 3D under the control of three stepper motors. In the experiment, the programmed trajectories of the tumor were all sinusoidal waves in three directions, with amplitudes in right/left (RL), anterior/posterior (AP), and superior/inferior (SI) directions of 5, 5, and 20 mm, respectively.

The 4D-CT dataset of the physical phantom was acquired using a Philips Brilliance 16 CT Scanner. The x-ray projections were acquired using Varian TrueBeam On-Board Imaging (OBI) system. The x-ray imager has a resolution of 1024×768 pixels with a physical size of 40×30 cm². The projections were downsampled to 512×384 pixels. We first collected 659 projections in half-fan mode with 125 kVp, 60 mA, and 20 ms exposure time and a CBCT image was reconstructed using a FDK algorithm.⁴³ The purpose of this CBCT image was to allow for rigid registration with the averaged 4D-CT images. After that, projections were acquired in Dynamic Gain Fluoro mode with 125 kVp, 60 mA, and 20 ms exposure time at a frequency of 11 frames/s. The projection angle was at 0° . The projections at the first 5 s were used for intensity baseline corrections, and the rest were used to test our method.

FIG. 4. [(a)–(c)] The 1st, 2nd, 3rd PCA coefficient results. [(d)–(f)] The residual error of the 1st, 2nd, 3rd PCA coefficient results compared with the ground truth.

TABLE I. Errors in the predicted PCA coefficients in different cases.

Number	Cases		PCA coefficient error		
	Sparse training	Intensity correction	First	Second	Third
1	Yes	No	0.27	3.55	3.08
2	No	Yes	0.10	0.11	0.22
3	No	No	0.31	3.03	2.89
4	Yes	Yes	0.05	0.14	0.38

We also expanded our experimental study using the breathing curve of one real patient. The tumor trajectory was measured on a Cyberknife Synchrony Respiratory Motion Tracking System. Its length was ~ 200 s. The SI component of the tumor trajectory was input to our phantom to drive the tumor motion in the SI direction. Motion amplitudes in the other two directions were found to be much smaller and hence they were neglected in our study. Once we set the phantom to motion, a 4D-CT data acquisition started. After that, the phantom was transferred to the OBI system, on which X-ray projections were acquired. All the parameters in this part are identical to those in the previous experiment.

3. RESULTS

3.A. Synthetic patient case

3.A.1. Sparse learning results

Figure 3 shows the results of sparse learning. We used partitions indicated in Fig. 3 with a square size of $4.97 \times 4.97 \text{ cm}^2$. (The reason for this choice will be presented in Sec. 3.C.2.) After running the sparse learning algorithm using the ten 4D-CT breathing cycle images, it was found that the patches indicated by the red rectangles are automatically selected and a matrix X^* is determined at the same time to map their intensities into the PCA coefficients of the DVFs. From this figure, we can see that our algorithm selects the patches that probably contain the richest motion information among all the patches, namely, those around the diaphragm. This result reflects the efficacy of our algorithm to a certain extent.

3.A.2. Motion prediction results: PCA coefficients

After sparse learning, we can use the built model to predict the motion for each input projection image and hence generate a corresponding volumetric image. We first show in Figs. 4(a)–4(c) the 1st, 2nd, and 3rd calculated PCA coefficients at 60 time points within a breathing cycle using the corresponding projections. In this synthetic case, the ground truth values are available, and hence we also present in Figs. 4(d)–4(f) residual errors. Quantitatively, Table I presents the errors in different cases, where the error is defined as $e = \|w - w_0\| / \|w_0\|$. Here, w is a vector consisting of the PCA coefficients of interest, and w_0 is the ground truth one. $\|\cdot\|$ denotes standard L_2 norm. Based on the figures and the table, we have the following observations. (1) Comparing the results with and without intensity correction, namely, rows 1 and 3

versus rows 2 and 4, it is found that there is a large error in the prediction results, if the intensity is not corrected. This implies the importance of the intensity correction step. In fact, because our model predicts the PCA coefficients directly using intensity values, any error in the intensity propagates to the prediction results. Therefore, an intensity correction step is needed to avoid systematic errors. The same behavior is also observed in other experiments in this paper. Hence, in the rest of the paper, we will always perform the intensity correction step. (2) After the gross prediction error due to intensity error is removed, we can observe the efficacy of sparse learning by comparing row 2 and 4. It was found that our model is able to reduce the prediction error of the 1st PCA coefficients. Although the error in the second and the third PCA components is slightly bigger, since the motion is dominated by the first component, the overall results in DVF will be better, as will be shown later. (3) There are actually two error sources, namely, intensity inaccuracy and prediction model inaccuracy. In this case, the first one dominates, and hence the similar amount of error in row 1 and 3 is observed, when the intensity is left uncorrected.

3.A.3. Motion prediction results: Images

The 60 volumetric images can be accordingly generated using the predicted motion vector fields. We compare these images with the corresponding ground truth ones. Since each test image is generated by deforming the reference image and the error in the predicted DVF is inevitable, it is expected that there will be structural location inaccuracy. We quantify the comparison using a generalized gamma index⁴⁴ in this context, which considers differences between two images with both

FIG. 5. The gamma index result of predicted images in one breath cycle.

TABLE II. Errors of each voxel in vector field. (Result with sparse learning/result without sparse learning.)

	$e_{\max}$ (mm)	$\bar{e}$ (mm)	$e_{95\%}$ (mm)
SI	2.12/3.47	0.08/0.14	0.64/1.80
RL	0.87/1.22	0.03/0.08	0.37/1.14
AP	1.07/1.47	0.05/0.13	0.45/1.21
Amplitude	2.22/3.74	0.09/0.23	0.92/2.40

intensity difference and spatial difference. Specifically,

$$\gamma(x) = \min_y \sqrt{\left[\frac{f_0(x) - f(y)}{\text{IC}} \right]^2 + \left[\frac{x - y}{\text{DC}} \right]^2}, \quad (8)$$

where f and f_0 are the test and the ground truth images. IC and DC are intensity and deformation criteria for the test, which are chosen as 10 HU and 0.5 mm, respectively. It is understood that a voxel at x passes the test, if $\gamma(x) < 1$. We report the passing rate of the whole image, i.e., the percentage of the voxels that pass the test.

Figure 5 shows the gamma index results of the predicted images in one breath cycle. It is clear that the proposed sparse learning model outperforms the other one, in that the gamma test passing rate is consistently higher. We further examined the improvement in vector field result quantitatively. Maximum error $e_{\max}$, 95th percentile error $e_{95\%}$, and mean error $\bar{e}$ are reported in Table II for the SI, RL, and AP components of the vector field, as well as for the amplitude. It is observed that the sparse learning method can greatly reduce the error at all items.

3.B. Real phantom case

Figure 6(a) is a coronal view of the 4D-CT at one breathing phase. Several subfigures in Fig. 6(b) are zoom-in views of the

tumor area in the images reconstructed by our method, which demonstrate tumor motion. Note that the tumor area is blurred in all the reconstructed images. This is ascribed to the blurring artifact seen in the 4D-CT images. As the reconstructed images are obtained by deforming a reference image with the reconstructed vector fields, the blurring artifacts are carried over.

For this real phantom case, the ground truth data for the PCA coefficients, DVFs, and the images were not available. In fact, although we can obtain the CT of a static phantom, we do not have the volumetric image at each moment, once it is set to move. While it is possible to use 4D-CT instead to obtain images at several breathing phases and interpolate to other time point, the resulting images may not be accurate and probably cannot be regarded as ground truth any more. However, the tumor motion trajectory is known, as it was controlled by the step motors. To derive the motion trajectory in the reconstructed images, we segmented the tumor in the reconstructed images using a simple threshold method, and the centroid of the tumor was calculated. In Fig. 6(c), the 3D tumor trajectories obtained with and without using the sparse learning method are compared with the ground truth, and the sparse learning method is seen to yield a trajectory closer to the ground truth. In addition, we have also plotted the motion coordinates along the SI, RL, and AP directions as functions of the breathing phase in Figs. 6(d)–6(f). Again, the sparse learning method yields better agreement with the ground truth. To further quantify the improvement, maximum error, 95-percentile error, and mean error along the SI, RL, and AP directions, as well as the magnitude, are summarized in Table III. It is observed that the sparse learning method can greatly reduce errors and hence increase tumor location accuracy. In particular, the mean error can be reduced by our proposed method to 0.82 mm, compared to the 1.66 mm error without using sparse learning.

Figure 7 shows the results for the phantom case but driven by a realistic tumor motion curve. Figure 7(a) is the curve

FIG. 6. (a) Coronal view of one phase of 4D-CT, display window (−1 000 200); (b) coronal view at different phases of the generated images from the proposed method; [(c)–(f)] The tumor center's trajectory, and SI, RL, and AP directions, respectively.

TABLE III. Errors in tumor localization in the phantom case. (Result after sparse learning/result without sparse learning.)

	$e_{\max}$ (mm)	$\bar{e}$ (mm)	$e_{95\%}$ (mm)
SI	1.69/3.26	0.66/1.39	1.42/3.11
RL	0.67/1.01	0.31/0.51	0.61/0.89
AP	0.69/0.79	0.27/0.54	0.56/0.77
Amplitude	1.74/3.32	0.82/1.66	1.55/3.14

along the SI direction. There clearly exists motion irregularity, e.g., amplitude and periodic variations and base-line shift. In Fig. 7(b), we compared the predicted tumor center location with the ground truth one in the time interval between 165 and 190 s. Each red circle corresponds to one x-ray projection, which was acquired at ~ 10 Hz on OBI. We found that for the majority part of the motion, our method was able to predict it to a good extent. In particular, the breathing period was slightly different in this interval of interest from that at the beginning portion used in 4D-CT scans. Our method can capture this variation of periodicity. For the amplitude variation, however, we observed inaccurate results at the region around extrema of the motion range. This indicated that the capability of our model to address amplitude irregularity should be improved in future studies. Quantitatively, maximum error and mean error of the predicted tumor center location were 1.4 and 0.6 mm, respectively, which are similar to those in the previous phantom case with a regular tumor motion pattern.

3.C. Robustness analysis

In this section, we would like to examine how the proposed sparse learning method is robust against different parameters.

FIG. 7. (a) SI motion curve of one real patient. Dashed box indicates the time interval from 165 to 190 s used to test our method. (b) The predicted tumor center compared with ground truth in this interval.

FIG. 8. Gamma index result as the function of patches number selected by sparse learning. The patches indicated by the rectangles are selected by sparse learning algorithm.

3.C.1. Effects of sparse level

It is necessary to balance the fidelity term and the sparse term in Eq. (4) by adjusting the parameter λ . It is interesting to see how this weighting factor affects the results; by tuning the parameter, the number of selected patches varies. As shown in Fig. 8, we calculated the averaged gamma index passing rate at different degrees of sparsity. The corresponding selected patches were also plotted. The results show that a high passing rate above 97% can be achieved using 4–13 patches. We also plotted the mean error of the calculated DVF, which attains minimum values in this regime. When the algorithm overemphasizes sparseness such that only one patch is selected, the patch does not contain enough motion information and hence the passing rate is low. In contrast, when numerous patches are selected, the model approaches the one without sparse learning, yielding again a low passing rate. It is interesting to observe that, regardless of the number of patches selected, there are some key locations that are

FIG. 9. The gamma index result as a function of patch sizes.

FIG. 10. Patch selection results with diaphragm removed. In (a) and (b), the patches indicated by the black rectangles are selected by sparse learning algorithm.

always chosen by the algorithm, e.g., those around the diaphragm area. This also demonstrates the accuracy of our patch selection algorithm to a certain extent.

3.C.2. Effects of patch size

We have also studied the impact of patch sizes. For each given combination u, v in the unit of pixel number, we conducted the training process and then used the trained model to generate a set of images. The gamma passing rate under each u, v is shown in grayscale in Fig. 9. It is observed that there exists an optimal patch size of 64×64 pixel² under which the highest passing rate is achieved. This is the size we used in previous studies. When the patch is too small, the selected patches do not contain sufficient motion information, as the key features that reflect motion, e.g., edges, sometimes move out of the selected patch. In contrast, a large patch with too many pixels dilutes the motion information contained in it. In either case, the passing rate is relatively low.

3.C.3. Impact of diaphragm

Although it has been demonstrated that the patch selection algorithm can automatically choose patches with motion information, as seen in Figs. 3 and 8, the diaphragm was a very clear motion feature in the projection images, which made the selection easy in those cases. We wanted to examine whether the algorithm was robust in the absence of the diaphragm. As such, we masked-out the entire diaphragm area in the projections, and conducted the study again. The results are in Fig. 10. It was found that the patches selected by sparse learning are divided into two parts, with the first three patches in the wall chest area, and the other two in the area near the diaphragm. Figure 11 shows the gamma index results of the predicted images in one breath cycle. The results show that the proposed sparse learning model again outperforms the one without sparse learning. Yet excluding the diaphragm area definitely makes the problem harder, which is clearly indicated by the reduced passing rates in Fig. 11 compared with those in Fig. 5 for both cases. Figure 11 also shows that the improvements of sparse learning are more obvious in the absence of the diaphragm. Within this context, the sparse learning model can still identify patches with important motion information and use them exclusively

to predict lung motion. In contrast, the model without sparse learning uses all the patches. Since motion information is weak, it is easily impeded by other irrelevant information, e.g., noise.

3.D. Computation time

Finally, we report the computation time for our algorithm in both the simulation and the experimental cases. Specifically, the time spent on training the model is broken into five parts for (1) deformable image registration, (2) PCA of the vector field, (3) simulating projections via DRR, (4) sparse learning, and (5) intensity correction. Once the model is built, it is applied to a projection to yield a volumetric image. The time on this step (6) is also reported. Of these, steps (1) and (3) are performed on an Nvidia Titan GPU card using our previously developed packages.^{35,42} Step (6) is also performed on a GPU card as it contains simply matrix–vector operations and volume interpolations, which are GPU-friendly. Other steps are conducted on a CPU platform in MATLAB. Detailed computation time is shown in Table IV. The time spent on the model training part is relatively long due to the many complicated computations. In particular, the deformable image registration step dominates the calculation time, as ten volumetric images are registered. PCA and sparse learning are executed in MATLAB, which can be further sped up with optimal implementations. The DRR part is also long, as polyenergetic projections are computed in our study. It will be our future

FIG. 11. The gamma index result of prediction phase image in one breath cycle in the absence of the diaphragm.

TABLE IV. Computation time of our method.

Case	Model training (s)					Generate a volumetric image (s)
	Registration	PCA	DRR	Sparse learning	Intensity correction	
Simulation	78.3	5.28	8.26	18.9	0.14	0.09
Experiment	75.5	5.15	8.26	18.7	0.14	0.09

research to accelerate computations in this model training part. In contrast, once a model is built, it takes only 0.09 s to predict a volumetric image on GPU based on a projection image, which implies the feasibility of using our model in online applications.

4. CONCLUSION AND DISCUSSIONS

In this paper, we have presented a method aimed at solving two problems in our previously developed algorithm for single projection-based volumetric image reconstruction, namely, robustness and computational efficiency. As opposed to obtaining a deformation vector field via forward projection matching, a sparse learning approach was employed. To fully extract motion information hidden in projection images, we partitioned a projection into small patches. At the training stage, we utilized sparse learning to automatically select patches that have a high correlation with the PCA coefficients of the lung motion model. A motion model was built at this stage together with the patch selection. We have also proposed an intensity baseline correction method based on the partitioned projection, in which the first and the second moments of pixel intensities at a patch in a simulated image were matched with those in a measured image via a linear transformation. The proposed method has been validated in simulated data and real phantom data. It has been found that the algorithm is able to identify patches that contain relevant motion information, specifically the diaphragm region. An intensity correction step is important to remove the systematic error in motion prediction. After that, each intensity-corrected projection can be used to generate a volumetric image. For the simulation case, the sparse learning model yields reduced prediction error in PCA coefficients and a higher gamma passing rate in the reconstructed image compared to the model without sparse learning but with intensity correction. In the phantom case, the predicted tumor motion trajectory was successfully reconstructed with 0.82 mm mean error in 3D space for the regular motion trajectory and 0.6 mm, when a real motion trajectory of a patient was used. The algorithm robustness with respect to sparse level, patch size, and existence of diaphragm was also tested.

Generally speaking, all the information in an image should be used for model training and application. The question here is how to build a predictive model that relates pixel intensities D with the PCA coefficients W of the DVFs. In our approach, we assumed a linear model of $XD = W$. Given a set of training data D and W , we would like to train the system to derive a system matrix X . A simple and straightforward approach would

be to solve an optimization problem $X = \operatorname{argmin}_X |XD - W|^2$. This approach uses all the information in a projection to train the model and more importantly, it treats all the pixel information with equal importance. However, it is expected that different portions of a projection contain different amounts of motion information. Since errors are inevitable on the images, using those parts that contain little or no motion information, but relatively a large amount of error, could mislead the model determination. Hence, a second approach is to solve an optimization problem $X = \operatorname{argmin}_X |XD - W|_A^2$, where the weighted L_2 norm specified by a matrix A adjusts the relative importance of different parts in an image. Yet this matrix A has to be specified empirically or determined via a certain algorithm. Our method actually belongs to the second category. However, we took a simple assumption that the weights for pixels in a patch are binary, namely, we either include or exclude a patch in the training process. More importantly, the determination of this simplified weighting matrix A is achieved via a sparse learning algorithm, together with the determination of the system matrix X . After that, the matrix X can be used to predict PCA coefficients given projection data D . It is our approach via binary weights that makes the multiplication of XD only involving those selected patches on the projection image.

Another issue is the impact of scattered photons from the MV beam to the kV detector, when our method is actually used during treatment delivery. The amount of scatter depends on patient geometry, dose rate, and field shape. For lung cancer patient, it is probably not high due to the low electron density of the lung area. For instance, in the experiments performed in Van's paper,⁴⁵ the impact of MV scatter to the CBCT in a lung cancer patient case was found small. However, it is our future study through experiments to determine how severe this problem is. If so, the intensity correction method developed in this paper is probably sufficient to address this issue. In fact, the main reason causing intensity mismatch between the measured and the calculated projection images in our algorithm was kV photon scatter. Our developed method was shown effective in terms of removing its impact and maintaining model accuracy. As the kV and the MV scatter signals on the detector are expected to be similar, namely, a low-frequency background one, we expect our algorithm is able to remove it to a satisfactory extent. Furthermore, if additional work were found necessary for MV scatter removal, mature algorithms^{45,46} developed over the years could be utilized in our system.

We would also like to discuss a few future directions. First, real patient studies are needed. Considering the complexity of patient respiratory motion, e.g., amplitude/phase irregularities

and baseline drift, only comprehensive tests in patient cases can evaluate the performance of the algorithm and hence its applicability in real clinical settings. Second, this current work only dealt with the case of a single fixed projection angle, with the potential application of reconstructing volumetric images during IMRT beam delivery. When it comes to volumetric modulated arc therapy (VMAT), the beam rotates continuously around a patient. A model built for a particular projection angle will be invalid for other beam angles. A potential solution could be to build a number of additional models, one for each projection angle. Taking into considerations of model correlation among neighboring angles may also improve model accuracy. Yet this brute force approach would certainly reduce computational efficiency. It will also be our future research topic to develop novel methods to extend the current algorithm for use with VMAT.

ACKNOWLEDGMENT

This work is supported in part by NIH (Nos. 1R01CA154747-01, 1R21CA178787-01A1, and 1R21EB017978-01A1), National Natural Science Foundation of China (Nos. 81301940 and 81428019), and Guangdong Strategic emerging industry core technology research (No. 2011A081402003). The preliminary results of this work were presented in the 2014 AAPM Annual meeting as a "Robust real-time volumetric imaging based on one single projection" abstract (Yuan Xu *et al.* 2014). The authors would like to thank Ms. Dee Hill for proof-reading this manuscript.

^{a)}Electronic addresses: xun.jia@utsouthwestern.edu; steve.jiang@utsouthwestern.edu; and smart@smu.edu.cn

¹S. B. Jiang, "Radiotherapy of mobile tumors," presented at the Seminars in Radiation Oncology (2006).

²P. J. Keall, G. S. Mageras, J. M. Balter, R. S. Emery, K. M. Forster, S. B. Jiang, J. M. Kapatoes, D. A. Low, M. J. Murphy, B. R. Murray, C. R. Ramsey, M. B. Van herk, S. S. Vedam, J. W. Wong, and E. Yorke, "The management of respiratory motion in radiation oncology report of AAPM task group 76," *Med. Phys.* **33**, 3874–3900 (2006).

³T. Pan, T.-Y. Lee, E. Rietzel, and G. T. Chen, "4D-CT imaging of a volume influenced by respiratory motion on multi-slice CT," *Med. Phys.* **31**, 333–340 (2004).

⁴P. Keall, G. Starkschall, H. E. E. Shukla, K. Forster, V. Ortiz, C. Stevens, S. Vedam, R. George, T. Guerrero, and R. Mohan, "Acquiring 4D thoracic CT scans using a multislice helical method," *Phys. Med. Biol.* **49**, 2053–2067 (2004).

⁵T. Pan, X. Sun, and D. Luo, "Improvement of the cine-CT based 4D-CT imaging," *Med. Phys.* **34**, 4499–4503 (2007).

⁶R. Li, J. H. Lewis, L. I. Cervino, and S. B. Jiang, "4D CT sorting based on patient internal anatomy," *Phys. Med. Biol.* **54**, 4821–4833 (2009).

⁷T. Li, A. Koong, and L. Xing, "Enhanced 4D cone-beam CT with inter-phase motion model," *Med. Phys.* **34**, 3688–3695 (2007).

⁸G.-H. Chen, J. Tang, and S. Leng, "Prior image constrained compressed sensing (PICCS): A method to accurately reconstruct dynamic CT images from highly undersampled projection data sets," *Med. Phys.* **35**, 660–663 (2008).

⁹S. Leng, J. Zambelli, R. Tolakanahalli, B. Nett, P. Munro, J. Star-Lack, B. Paliwal, and G.-H. Chen, "Streaking artifacts reduction in four-dimensional cone-beam computed tomography," *Med. Phys.* **35**, 4649–4659 (2008).

¹⁰S. Rit, J. W. Wolthaus, M. van herk, and J.-J. Sonke, "On-the-fly motion-compensated cone-beam CT using an a priori model of the respiratory motion," *Med. Phys.* **36**, 2283–2296 (2009).

¹¹M. Ahmad, P. Balter, and T. Pan, "Four-dimensional volume-of-interest reconstruction for cone-beam computed tomography-guided radiation therapy," *Med. Phys.* **38**, 5646–5656 (2011).

¹²J. C. Park, S. H. Park, J. H. Kim, S. M. Yoon, S. S. Kim, J. S. Kim, Z. Liu, T. Watkins, and W. Y. Song, "Four-dimensional cone-beam computed tomography and digital tomosynthesis reconstructions using respiratory signals extracted from transcutaneously inserted metal markers for liver SBRT," *Med. Phys.* **38**, 1028–1036 (2011).

¹³Z. Zheng, M. Sun, J. Pavkovich, and J. Star-Lack, "Fast 4D cone-beam reconstruction using the McKinnon-Bates algorithm with truncation correction and nonlinear filtering," *Proc. SPIE* **7961**, Medical Imaging 2011: Physics of Medical Imaging, 79612U (2011).

¹⁴X. Jia, Z. Tian, Y. Lou, J.-J. Sonke, and S. B. Jiang, "Four-dimensional cone beam CT reconstruction and enhancement using a temporal nonlocal means method," *Med. Phys.* **39**, 5592–5602 (2012).

¹⁵L. Ren, I. J. Chetty, J. Zhang, J.-Y. Jin, Q. J. Wu, H. Yan, D. M. Brizel, W. R. Lee, B. Movsas, and F.-F. Yin, "Development and clinical evaluation of a three-dimensional cone-beam computed tomography estimation method using a deformation field map," *Int. J. Radiat. Oncol., Biol., Phys.* **82**, 1584–1593 (2012).

¹⁶H. Gao, R. Li, Y. Lin, and L. Xing, "4D cone beam CT via spatiotemporal tensor framelet," *Med. Phys.* **39**, 6943–6946 (2012).

¹⁷J.-F. Cai, X. Jia, H. Gao, S. B. Jiang, Z. Shen, and H. Zhao, "Cine cone beam CT reconstruction using low-rank matrix factorization: Algorithm and a proof-of-principle study," preprint [arXiv:1204.3595](https://arxiv.org/abs/1204.3595) (2012).

¹⁸J. Wang and X. Gu, "High-quality four-dimensional cone-beam CT by deforming prior images," *Phys. Med. Biol.* **58**, 231–246 (2013).

¹⁹H. Yan, X. Zhen, M. Folkerts, Y. Li, T. Pan, L. Cervino, S. B. Jiang, and X. Jia, "A hybrid reconstruction algorithm for fast and accurate 4D cone-beam CT imaging," *Med. Phys.* **41**, 071903 (12pp.) (2014).

²⁰R. Zeng, J. A. Fessler, and J. M. Balter, "Respiratory motion estimation from slowly rotating x-ray projections: Theory and simulation," *Med. Phys.* **32**, 984–991 (2005).

²¹D. A. Low, P. J. Parikh, W. Lu, J. F. Dempsey, S. H. Wahab, J. P. Hubenschmidt, M. M. Nystrom, M. Handoko, and J. D. Bradley, "Novel breathing motion model for radiotherapy," *Int. J. Radiat. Oncol., Biol., Phys.* **63**, 921–929 (2005).

²²Q. Zhang, A. Pevsner, A. Hertanto, Y.-C. Hu, K. E. Rosenzweig, C. C. Ling, and G. S. Mageras, "A patient-specific respiratory model of anatomical motion for radiation treatment planning," *Med. Phys.* **34**, 4772–4781 (2007).

²³R. Li, B. P. Fahimian, and L. Xing, "A Bayesian approach to real-time 3D tumor localization via monoscopic x-ray imaging during treatment delivery," *Med. Phys.* **38**, 4205–4214 (2011).

²⁴R. Li, J. H. Lewis, X. Jia, X. Gu, M. Folkerts, C. Men, W. Y. Song, and S. B. Jiang, "3D tumor localization through real-time volumetric x-ray imaging for lung cancer radiotherapy," *Med. Phys.* **38**, 2783–2794 (2011).

²⁵R. Li, J. H. Lewis, X. Jia, T. Zhao, W. Liu, S. Wuenschel, J. Lamb, D. Yang, D. A. Low, and S. B. Jiang, "On a PCA-based lung motion model," *Phys. Med. Biol.* **56**, 6009–6030 (2011).

²⁶C.-R. Chou, B. Frederick, X. Liu, G. Mageras, S. Chang, and S. Pizer, "Claret: A fast deformable registration method applied to lung radiation therapy," presented at the Fourth International (MICCAI) Workshop on Pulmonary Image Analysis (2011).

²⁷D. Staub, A. Docef, R. S. Brock, C. Vaman, and M. J. Murphy, "4D cone-beam CT reconstruction using a motion model based on principal component analysis," *Med. Phys.* **38**, 6697–6709 (2011).

²⁸C.-R. Chou, B. Frederick, G. Mageras, S. Chang, and S. Pizer, "2D/3D image registration using regression learning," *Comput. Vision Image Understanding* **117**, 1095–1106 (2013).

²⁹Y. Zhang, J. Yang, L. Zhang, P. A. Balter, and L. Dong, "Modeling respiratory motion for reducing motion artifacts in 4D CT images," *Med. Phys.* **40**, 041716 (13pp.) (2013).

³⁰C. A. McBain, A. M. Henry, J. Sykes, A. Amer, T. Marchant, C. M. Moore, J. Davies, J. Stratford, C. McCarthy, and B. Porritt, "X-ray volumetric imaging in image-guided radiotherapy: The new standard in on-treatment imaging," *Int. J. Radiat. Oncol., Biol., Phys.* **64**, 625–634 (2006).

³¹D. A. Jaffray and J. H. Siewerdsen, "Cone-beam computed tomography with a flat-panel imager: Initial performance characterization," *Med. Phys.* **27**, 1311–1323 (2000).

³²D. A. Jaffray, J. H. Siewerdsen, J. W. Wong, and A. A. Martinez, "Flat-panel cone-beam computed tomography for image-guided radiation therapy," *Int. J. Radiat. Oncol., Biol., Phys.* **53**, 1337–1349 (2002).

- ³³J. McClelland, D. J. Hawkes, T. Schaeffter, and A. P. King, "Respiratory motion models: A review," *Med. Image Anal.* **17**, 19–42 (2013).
- ³⁴M. Söhn, M. Birkner, D. Yan, and M. Alber, "Modelling individual geometric variation based on dominant eigenmodes of organ deformation: Implementation and evaluation," *Phys. Med. Biol.* **50**, 5893–5908 (2005).
- ³⁵X. Gu, H. Pan, Y. Liang, R. Castillo, D. Yang, D. Choi, E. Castillo, A. Majumdar, T. Guerrero, and S. B. Jiang, "Implementation and evaluation of various demons deformable image registration algorithms on a GPU," *Phys. Med. Biol.* **55**, 207–219 (2010).
- ³⁶R. L. Siddon, "Fast calculation of the exact radiological path for a 3-dimensional CT array," *Med. Phys.* **12**, 252–255 (1985).
- ³⁷M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables," *J. R. Stat. Soc. Ser. B-Stat. Methodol.* **68**, 49–67 (2006).
- ³⁸X. Jia, C. Men, Y. Lou, and S. B. Jiang, "Beam orientation optimization for intensity modulated radiation therapy using adaptive l2, l1-minimization," *Phys. Med. Biol.* **56**, 6205–6222 (2011).
- ³⁹B. Dong, Y. J. Graves, X. Jia, and S. B. Jiang, "Optimal surface marker locations for tumor motion estimation in lung cancer radiotherapy," *Phys. Med. Biol.* **57**, 8201–8215 (2012).
- ⁴⁰J. Liu, S. Ji, and J. Ye, *SLEP: Sparse Learning With Efficient Projections* (Arizona State University, Tempe, AZ, 2009).
- ⁴¹X. Zhen, X. Gu, H. Yan, L. Zhou, X. Jia, and S. B. Jiang, "CT to cone-beam CT deformable registration with simultaneous intensity correction," *Phys. Med. Biol.* **57**, 6807–6826 (2012).
- ⁴²X. Jia, H. Yan, L. Cervino, M. Folkerts, and S. B. Jiang, "A GPU tool for efficient, accurate, and realistic simulation of cone beam CT projections," *Med. Phys.* **39**, 7368–7378 (2012).
- ⁴³L. A. Feldkamp, L. C. Davis, and J. W. Kress, "Practical cone beam algorithm," *J. Opt. Soc. Am. A* **1**, 612–619 (1984).
- ⁴⁴D. A. Low, W. B. Harms, S. Mutic, and J. A. Purdy, "A technique for the quantitative evaluation of dose distributions," *Med. Phys.* **25**, 656–661 (1998).
- ⁴⁵M. van Herk, L. Ploeger, and J. J. Snoke, "A novel method for megavoltage scatter correction in cone-beam CT acquired concurrent with rotational irradiation," *Radiother. Oncol.* **100**, 365–369 (2011).
- ⁴⁶C. Ling, P. Zhang, T. Etmektzoglou, J. Star-Lack, M. Sun, E. Shapiro, and M. Hunt, "Acquisition of MV-scatter-free kilovoltage CBCT images during RapidArc or VMAT," *Radiother. Oncol.* **100**, 145–149 (2011).