Local Intensity Feature Tracking and Motion Modeling for Respiratory Signal Extraction in Cone Beam CT Projections

Salam Dhou, *Student Member, IEEE*, Yuichi Motai*, *Member, IEEE*, and Geoffrey D. Hugo

***Abstract*—Accounting for respiration motion during imaging can help improve targeting precision in radiation therapy. We propose local intensity feature tracking (LIFT), a novel markerless breath phase sorting method in cone beam computed tomography (CBCT) scan images. The contributions of this study are twofold. First, LIFT extracts the respiratory signal from the CBCT projections of the thorax depending only on tissue feature points that exhibit respiration. Second, the extracted respiratory signal is shown to correlate with standard respiration signals. LIFT extracts feature points in the first CBCT projection of a sequence and tracks those points in consecutive projections forming trajectories. Clustering is applied to select trajectories showing an oscillating behavior similar to the breath motion. Those "breathing" trajectories are used in a 3-D reconstruction approach to recover the 3-D motion of the lung which represents the respiratory signal. Experiments were conducted on datasets exhibiting regular and irregular breathing patterns. Results showed that LIFT-based respiratory signal correlates with the diaphragm position-based signal with an average phase shift of 1.68 projections as well as with the internal marker-based signal with an average phase shift of 1.78 projections. LIFT was able to detect the respiratory signal in all projections of all datasets.**

***Index Terms*—Cone beam computed tomography (CBCT), image motion analysis, respiration signal.**

I. INTRODUCTION

RESPIRATORY motion extraction from cone beam computed tomography (CBCT) images is an important task in biomedical engineering research [1]–[3]. It can be used to study the influence of organ motion on CBCT imaging [4]. Motion modeling can be used also for measuring the position and orientation of objects [5], [6] or for respiratory image sequence segmentation techniques [7]. Four-dimensional or respiration-correlated CT (4DCT) imaging techniques have become a basic task in radiation therapy planning. As the respiratory motion can be a major source of error in determining the position of thoracic and upper abdominal tumor targets during radiotherapy, extracting respiratory motion is a key task in reconstructing 4DCT. Volumetric image guidance techniques, such as 4D Cone Beam CT (4DCBCT), have been recently and rapidly integrated into the clinic for verifying tumor position during treatment and managing respiration-induced tissue motion [4], [8], [9]. An acquired respiratory signal serves as a surrogate for the tumor position. This surrogate is used to assign each projection to its appropriate breathing phase bin, in a process termed "sorting," prior to 4-D image reconstruction.

Respiratory signals can be extracted using external sources such as skin markers, abdominal belts, or spirometry [4], [10]. Those methods require additional equipment such as infrared cameras, detectors, or spirometry that may not be available. Another solution was an image-based measure of diaphragm position directly from the 4DCBCT radiographic projections [8], [11]. The extraction of a diaphragm position-based signal requires the diaphragm to be visible in all acquired CBCT projections, which is not possible in some commonly used CBCT systems that have limited longitudinal fields of view. Also, a number of studies have shown that tumor position is difficult to predict directly from the diaphragm or external surrogate motion, with reported errors of up to 6 mm in predicting tumor position from external marker position [12]–[21]. Another option is using transthoracically or bronchoscopically radio-opaque tumor markers that are implanted near the tumor and tracked to extract the respiratory motion [22], [23]. This method solves the problem, but the additional procedure to implant the markers is invasive and expensive, and any complications may significantly delay treatment initiation.

Thus, we propose local intensity feature tracking (LIFT), an image-based respiration signal extraction method. The contributions of this paper are twofold. First, the proposed method uses only the CBCT projections of the lung to extract the respiratory signal depending on tissue feature points local to the tumor. Second, the extracted respiration signal using LIFT correlates to the standard respiration signals. The specific novel contributions of this study are the following: 1) LIFT extracts the respiratory signal without dependence on a particular anatomical structure in the CBCT images (such as the diaphragm), and can be tuned to focus on a particular region of the anatomy (say, near a tumor).

Manuscript received May 25, 2012; revised September 9, 2012; accepted October 12, 2012. Date of publication November 12, 2012; date of current version January 16, 2013. This work was supported in part by the dean's office of the School of Engineering, the Presidential Research Incentive Fund at Virginia Commonwealth University, and National Science Foundation ECCS #1054333. *Asterisk indicates corresponding author.*

S. Dhou is with the Department of Electrical and Computer Engineering, Virginia Commonwealth University, Richmond, VA 23284 USA (e-mail: dhous@vcu.edu).

*Y. Motai is with the Deparment of Electrical and Computer Engineering, Virginia Commonwealth University, Richmond, VA 23284 USA (e-mail: ymotai@vcu.edu).

G. D. Hugo is with the Department of Radiation Oncology, Virginia Commonwealth University, Richmond, VA 23298 USA (e-mail: gdhugo@vcu.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TBME.2012.2226883

TABLE I
COMPARISON OF RESPIRATORY MOTION DETECTION METHODS

Method name	Procedure	Requirements
Diaphragm position[8],[11]	Measuring the position of the diaphragm in subsequent projections and using it as a surrogate to the respiration motion	Diaphragm should be visible in all projections
Internal Marker-based [22], [23]	Tracking the markers in subsequent projection images using image-based tracking techniques	Internal markers/ implanting procedure costs
Proposed LIFT	Tracking interest points local to the tumor position and 3D motion modeling of the 2D motion of the best selected motion trajectories	CT projections only

2) No other group has used 3-D motion modeling in recovering the 3-D motion of the lung and used it as a respiratory signal. LIFT works as follows. First, the CBCT projections of the entire dataset are grouped into arcs. Then, in each arc, we extract feature points in the lung. Those feature points are tracked from one projection to another using optical flow [24] to form trajectories of the feature points' correspondences. A selection criterion is applied to select the trajectories that show an oscillating behavior similar to respiration. Using those selected trajectories, the 3-D motion is recovered. Then, the 3-D rotation around the Z-axis of the patient represents the respiration motion in this study. CBCT projections are sorted into phases according to the respiration signal detected.

The advantage of LIFT over the external resources methods is that no external equipment is required. It also has the advantage over the diaphragm position method in that the diaphragm is not required to be visible in all CBCT projections. Moreover, no internal markers are required to be implanted as the signal is generated from tissue features local to the tumor position. We consider LIFT widely applicable as only the CBCT projections are required to generate the respiratory signal. There are several works related to our study. Bergner *et al.* [25] proposed a method for measuring motion between 4DCBCT projections based on dense optical flow using a Horn–Schunk [26] implementation. Their method was developed to improve the reconstruction quality in stationary regions of the anatomy. For respiration sorting, they used a diaphragm position-based method similar to [8]. Wachinger *et al.* [27] extract an image-based respiratory signal using manifold learning; however, this method was applied only for fixed view (fluoroscopic) images, as opposed to the rotational view (tomographic) projections used here.

The works in [11] and [28]–[30] are also related. In [11], thousands of interest points are selected and tracked across projections using a block matching algorithm. Signal processing techniques are then used to acquire a respiratory signal. In [28], Siochi developed a technique that locates a bounding box for the diaphragm motion for all projections based on two pairs of full-inhale and full-exhale views. However, the aforementioned methods [11], [28] require that the diaphragm be present in the field of view. In [29], the respiratory signal is acquired from an analysis of the variation in pixel values between projection images by developing a simple pixel value summation followed by a high-pass filtering. This method is effective, but its efficiency should be demonstrated for various CBCT patient cases and compared to other methods in the literature. Vergalasova *et al.* [30] proposed a markerless method for respiration signal detection based on the principles on the Fourier Transform theory. However, when the entire projections of patient data were used, the results showed a big phase shift between the extracted signal and the ground truth signal. Thus, LIFT is the first method to use feature tracking and 3-D motion modeling for respiratory signal extraction. Table I shows a comparison summary of respiration signal extraction methods.

The remainder of this paper is organized as follows. Section II describes in detail the proposed method. Section III presents the experimental results on four clinical datasets and a discussion of the results. Section IV concludes this paper.

II. PROPOSED RESPIRATORY MOTION EXTRACTION METHOD

Here, we present the proposed LIFT method. In Section II-A, we discuss feature extraction and tracking. In Section II-B, we present trajectory clustering. In Section II-C, respiratory motion detection is discussed. In Section II-D, projection phase sorting is presented.

A. Feature Extraction and Tracking

Feature points are extracted and tracked through projection sequences to detect the optical flow motion in the following three steps:

Step 1: We extract feature points at pixel locations equally spaced by a constant number of pixels. We do not extract feature points based on image intensity (like corners) because of the nature of transmission tomography in which corners visible in 2-D projection images may not correspond to actual high contrast boundaries in the 3-D anatomy. Also, choosing features at equally spaced locations allows the extraction of feature points in any CT projection image. An extracted feature point is represented by $p_{f,p} = (x_{f,p},\mathrm{y}_{f,p})$, where f is the projection number and p is the point number. We aim at analyzing the motion between any pair of consecutive projections. The inferior portion of projections containing the diaphragm was excluded to simulate common CBCT acquisition systems that have smaller longitudinal fields of view. Fig. 1 shows feature points extraction in one projection.

Step 2: Feature points are tracked through subsequent projection frames. Due to the respiration motion and different projection angles, a feature point has different locations (x- and y-coordinates) in every subsequent projection. Those locations of the feature point are called "correspondences." To find the displacement $\delta(\Delta x,\Delta y)$ that a point $p_{f,p}$ with intensity $I(x_{f,p}, y_{f,p}, \tau)$ makes when it moves from one image to the next, a single pixel cannot be tracked. That is because the value of the pixel can change due to noise, and be confused with adjacent pixels [31]. Thus, windows of pixels are tracked instead of single pixels. Since adjacent projections refer to the same scene taken

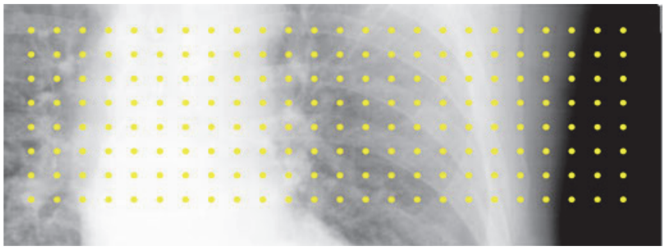


Fig. 1. Feature points extraction used in this study. Feature points are extracted at specific pixel locations spaced by a number of pixels (20 pixels in this figure). The inferior part of the projection image is excluded to simulate the CBCT systems that have a small field of view.

from slightly different viewpoints at τ and $\tau+\Delta\tau$, intensity patterns move in the image sequence satisfying the constraint property $I(x_{f,p}, \mathrm{y}_{f,p}, \tau) = I(x_{f,p} + \Delta x, \mathrm{y}_{f,p} + \Delta y, \tau + \Delta\tau)$. The next image can be defined as $J(p) = I(p - \delta) + \eta(p)$, where η is some noise. We find the displacement $\delta(\Delta x, \Delta y)$ that minimizes the sum of squared intensity differences between a past and current window w in the following cost function:

$$\varepsilon = \int_w [I(p-\delta) - J(p)]^2 dp \tag{1}$$

which can be written as $\varepsilon = \int_w (h - g\delta)^2 dp$, where $h = I(p) - J(p)$. The residue is minimized by differentiating ε with respect to δ and setting the result equal to zero as $\int_w (h - g\delta) g dA = 0$, where $(g\delta)g = (gg^T)\delta$, and δ is assumed to be constant within w. Therefore, we have $(\int_w gg^T dA)\delta = \int_w hgdA$. This system has two scalar equations and two unknowns which can be written as $G\delta = e$, where G can be computed from one image by estimating gradients and computing their second-order moments, and e can be computed from the difference between the two images along with the gradient computed previously. The displacement δ is then the solution of system [31].

Drift problem is a very classical issue when tracking feature points through long sequences. To overcome this problem, the quality of feature points is monitored during tracking by measuring the dissimilarity of the features between the first image in the sequence and the current image. The feature is abandoned when dissimilarity grows too large. Affine image changes are used to calculate dissimilarity as in [32].

Step 3: A trajectory is formed as sequence of point correspondences through F frames and is defined by t_p as $t_p = \{p_{1,p}, p_{2,p}, \ldots, p_{F,p}\}$. A set of trajectories is represented by $T = \{t_1, t_2, \ldots, t_P\}$, where P is the total number of trajectories which equals the total number of points. Trajectories are represented by a list of line-angle vectors (l,Θ), where l is the line between two adjacent feature points $p_{f,p}$ and $p_{f+1,p}$, and Θ is the value of the angle between two adjacent lines. The line-angle vector representation of a trajectory t_p of F point correspondences is $t_p = ((l_p^1, \Theta_p^1), \ldots, (l_p^{F-1}, \Theta_p^{F-2}))$. In a trajectory of F points, the number of lines l equals $F - 1$, while the number of angles Θ equals $F - 2$. This representation of trajectories is defined for the purpose of clustering. The details of two trajectories are shown in Fig. 2. Algorithm 1shows a summary of the steps used for feature extraction and tracking.

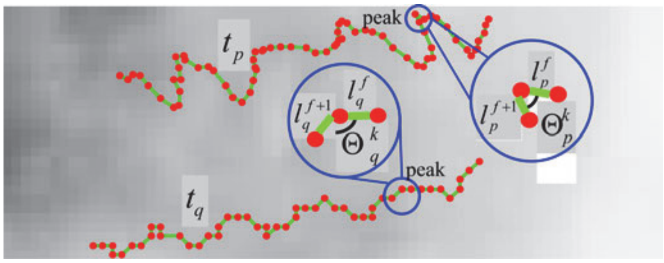
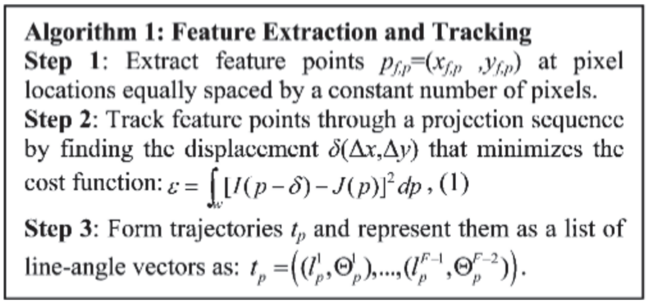
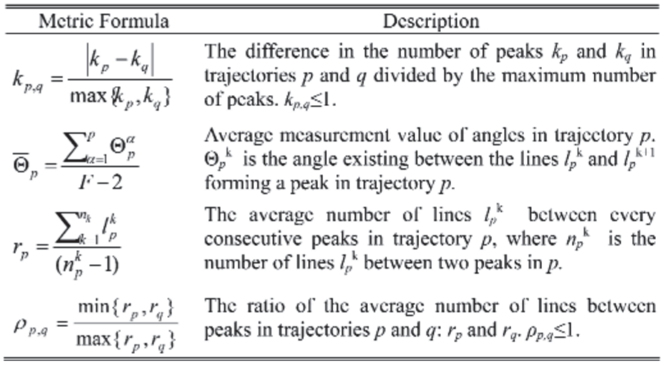


Fig. 2. Two trajectories formed by tracking a sequence of 47 frames. Both trajectories t_p and t_q are selected to be used in the reconstruction of the respiration signal. The first two segments of both trajectories are illustrated.



Algorithm 1: Feature Extraction and Tracking
Step 1: Extract feature points $p_{f,p}$=($x_{f,p}$,$y_{f,p}$) at pixel locations equally spaced by a constant number of pixels.
Step 2: Track feature points through a projection sequence by finding the displacement $\delta(\Delta x,\Delta y)$ that minimizes the cost function: $\varepsilon = \int_w [I(p-\delta) - J(p)]^2 dp$, (1)
Step 3: Form trajectories t_p and represent them as a list of line-angle vectors as: $t_p = \left((l_p^1, \Theta_p^1), \ldots, (l_p^{F-1}, \Theta_p^{F-2})\right)$.

TABLE II
TRAJECTORIES CLUSTERING METRICS



Metric Formula	Description
$k_{p,q} = \frac{\lvert k_p - k_q \rvert}{\max\{k_p, k_q\}}$	The difference in the number of peaks k_p and k_q in trajectories p and q divided by the maximum number of peaks. $k_{p,q} \leq 1$.
$\overline{\Theta}_p = \frac{\sum_{\alpha=1}^{p} \Theta_p^{\alpha}}{F-2}$	Average measurement value of angles in trajectory p. Θ_p^k is the angle existing between the lines l_p^k and l_p^{k+1} forming a peak in trajectory p.
$r_p = \frac{\sum_{k=1}^{n_k} l_p^k}{(n_p^k - 1)}$	The average number of lines l_p^k between every consecutive peaks in trajectory p, where n_p^k is the number of lines l_p^k between two peaks in p.
$\rho_{p,q} = \frac{\min\{r_p, r_q\}}{\max\{r_p, r_q\}}$	The ratio of the average number of lines between peaks in trajectories p and q: r_p and r_q. $\rho_{p,q} \leq 1$.

B. Shape-Based Trajectory Clustering

We cluster the trajectories resulting from tracking the feature points based on their motion behavior. Respiratory-induced motion in the thorax and upper abdomen is quasiperiodic and directed mainly in the superior–inferior direction (along the patient longitudinal axis). Thus, motion trajectories exhibiting this oscillating behavior may imply respiration motion (i.e., thoracic tissue areas). Other trajectories have only orbital motion due to the scanner rotation (i.e., bony areas). In order to detect a true respiration signal out of this mixed signal, we aim to minimize the effect of the orbital trajectories and use only the ones with the highest superior–inferior motion signals in the motion detection process. The following three steps show the process of trajectory clustering.

Step 1: A set of metrics for trajectory clustering is formulated to describe the shape of trajectories. Table II describes the set of metrics used and their formula. Those metrics compare trajectories based on their shapes using the number of peaks in each trajectory, angle measurements, and the average number of lines between peaks. A peak in a trajectory, as seen in Fig. 2, is the point that has the largest y-coordinate within a breath cycle (3–5 s). It is detected by traversing every trajectory to find the points with the largest y-coordinate within a breath cycle.

Those metrics are combined in one similarity measure to apply to every pair of trajectories. This similarity measure detects the similarity in the shape of trajectories and is insensitive to their location in the projection and length in pixels. The following is the definition of the similarity measure that uses the distance metrics defined in Table II

$$S(t_p,t_q) = d - \left(\begin{array}{l}\alpha k_{p,q} \\ + \beta \left|\bar{\Theta}_p - \bar{\Theta}_q\right| / \pi \\ + \psi(1-\rho_{p,q})\end{array}\right) \tag{2}$$

where t_p and t_q are two trajectories of the same length. d is the maximum similarity possible between any two trajectories. α, β, and Ψ are weights to adjust the significance of one metric to the other metrics. All metrics are represented as ratio or normalized difference. Each metric in (2) compares the pairs of trajectories based on one specific aspect of each trajectory's shape. The first metric compares trajectories based on the number of peaks. The second one compares the trajectories based on average angle measurement between the lines forming the peaks. The third one compares the trajectories based on the average number of lines between each consecutive peak. Each metric has its own influence on the result of similarity. The metrics are combined in a weighted summation to determine the overall similarity between each pair of trajectories. Those metrics are also weighted using three weighs (α, β, and Ψ) to increase or decrease the influence of a specific metric on the overall similarity.

When each of the weights α, β, and Ψ is set to 1, the maximum similarity possible d should be set to a numeric value that is equal to or greater than 3 in order to have a positive overall similarity $S(t_p,t_q)$. If d is set to 3, the overall similarity measure $S(t_p,t_q)$ will range from 0 (minimal similarity) to 3 (maximal similarity). The weights can be adjusted to give more significance to one metric than the other. For example, to emphasize on the average angle difference, such that similar trajectories should have very similar angles, β should be given a value greater than other weights α and Ψ, and the maximum similarity possible d should be changed accordingly.

Step 2: The hierarchical agglomerative clustering method [33], which is a type of hierarchical clustering, is used to classify trajectories. The similarity is computed for every pair of trajectories based on the similarity measure defined in (2). Clustering is achieved by first finding the closest pairs of trajectories and placing them into a cluster. The similarity between a single trajectory and the new cluster is computed as the average similarity between the single trajectory and all trajectories belonging to the cluster. Then, the most similar pair of trajectories/clusters is combined again in a new cluster until having two clusters eventually. One cluster contains "breath-like oscillating" trajectories and is denoted by T_b. The other one contains "orbital" trajectories and is denoted by T_o. This clustering method is meant to work on regular and irregular breathing patterns. Due to the irregular breathing, trajectories may have "abnormal" shapes. Since the whole anatomy is affected by the same regular/irregular respiration motion, most of the oscillating trajectories will have similar motion behaviors and thus similar shapes. So, they will be clustered in T_b, the cluster with breathing trajectories, and the other cluster T_o will contain either orbital or nonbreathing trajectories. Fig. 3 shows an example of trajectories clustering using our method.

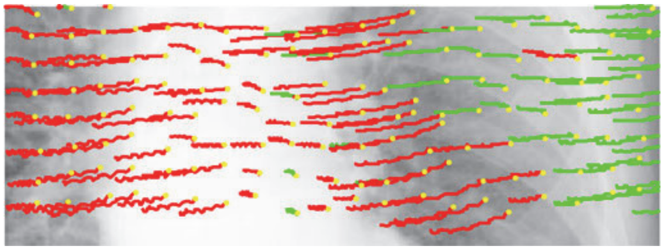


Fig. 3. Trajectories clustering in the projection images of range #250–300 of Patient 1 dataset. Red trajectories (cluster T_b) show a motion similar to the respiration motion, so they are selected to be used in the signal extraction. Green trajectories (cluster T_o) show an orbital motion, so they are discarded.

Step 3: The result of the clustering process has been evaluated after the process of clustering is done. Two validation criteria described in [34] are used in this study. The first criterion used is the compactness of a cluster which measures the average similarity of trajectories in a cluster. The compactness in cluster T_b is computed as

$$\bar{S}_b = \sum_{i=1}^{n_b} \sum_{j=i+1}^{n_b} S(t_i,t_j)/n_b \tag{3}$$

where n_b is the number of trajectories in T_b. The standard deviation between trajectories in the same cluster T_b is defined as

$$\sqrt{\sum_{i=1}^{n_b} \sum_{j=i+1}^{n_b} \left(S(t_i,t_j) - \bar{S}_b\right)^2 / n_b}.$$

The second criterion used is the isolation. It measures the separation of the two clusters by estimating the highest similarity to a trajectory outside the cluster. The isolation of the two clusters T_b and T_o is defined as

$$D(T_b,T_o) = \max(S(t_{i,b},t_{j,o})), \quad \forall i=1,\ldots,n_b, \quad j=1,\ldots,n_o. \tag{4}$$

The smaller the similarity between clusters, the greater the isolation. Algorithm 2 summarizes the steps of clustering.

C. Respiratory Motion Detection

For the detection of respiratory motion, we use 3-D motion modeling of the selected trajectories. We follow the following three steps to detect the respiratory motion.

Step 1: We group the subsequent projections into overlapping arcs. An *arc* is defined as a sequence of projections captured from a unique and continuous record of X-ray radiation. Projections are grouped into overlapping arcs because corresponding feature points are not visible in all projections due to scanner rotation. Overlapping the arcs allows for a breathing signal to

Algorithm 2: Shape-based Trajectory Clustering
Step 1: Formulate a similarity measure and apply to every pair of trajectories as in (2).
Step 2: Use hierarchical agglomerative clustering based on (2) to cluster trajectories into two clusters T_b and T_o.
Step 3: Validate clustering by computing compactness of T_b as in (3), and isolation of T_b and T_o as in (4).

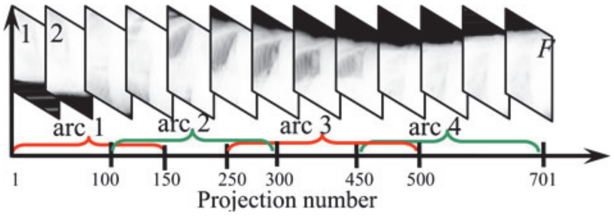


Fig. 4. Grouping consecutive projection images into arcs. Patient # 1 dataset images grouped into four arcs. Arc #1: 100–150, Arc #2: 100–300, Arc #3: 250–500, Arc #4: 450–701. The process is done manually.

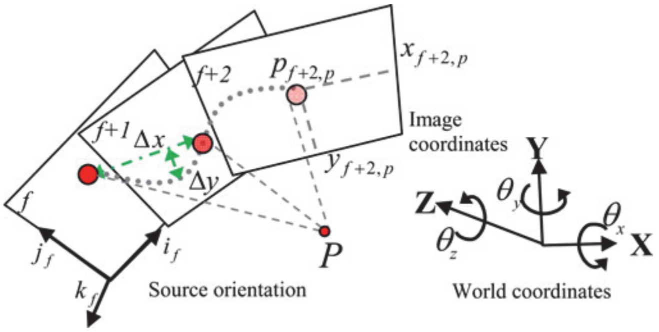


Fig. 5. Coordinate systems used in this method and one feature point tracked in three frames f, $f{+}1$, and $f{+}2$. Real world frame is represented by *XYZ* coordinates and CBCT projection frame is represented by *xy*. Source orientation is represented by the orientation vectors i_f, j_f, and k_f.

be correlated between two arcs in this region without having the same feature points in both arcs. The process of grouping projections into arcs is done manually based on their projection angle, regardless of their contents and/or the breathing signal they may carry. Projection grouping has no relation to the regularity or periodicity of breathing. The goal of grouping projections into arcs is to overcome the problem of feature points that become invisible because of the scanner rotation regardless of the breathing status. Fig. 4 shows dataset groupings into overlapping arcs.

Step 2: We detect the 3-D motion of the lung in each arc using the selected trajectories T_b clustered in Section II-B. Fig. 5 shows the coordinate systems used in this study. This figure shows one 2-D point tracking in three consecutive projection frames. The x-coordinate displacement Δx corresponds to the displacement caused by the orbital motion. Δy corresponds to the craniocaudal (up-down) position of the lung. To detect the 3-D motion of the lung, we use the structure-from-motion technique [35]. The 2-D points $p_{f,p} = (x_{f,p}, y_{f,p})$, of the set of trajectories T_b are filled in the measurement matrix $W: 2F \times P$ as in (5). Then, the mean m_f and n_f for each measurement type is subtracted off from W to yield the registered measurement matrix W^* as

$$W = \begin{bmatrix} x_{1,1} & \cdots & x_{1,P} \\ \vdots & \ddots & \vdots \\ x_{F,1} & \cdots & x_{F,P} \\ \hline y_{1,1} & \cdots & y_{1,P} \\ \vdots & \ddots & \vdots \\ y_{F,1} & \cdots & y_{F,P} \end{bmatrix} = \begin{bmatrix} x_{f,p} \\ y_{f,p} \end{bmatrix} \tag{5}$$

$$W^* = \begin{bmatrix} x_{f,p} \\ y_{f,p} \end{bmatrix} - \begin{bmatrix} m_f \\ n_f \end{bmatrix}. \tag{6}$$

To recover the 3-D motion, the registered measurement matrix W^* is decomposed by (SVD) into $W^* = U'D'V'^T$. The respective motion matrix M' is extracted as $M' = U'\sqrt{D'}$ and the true value of the motion matrix M: $2F \times 3$ is recovered as $M = M'\Lambda$, where Λ is a 3×3 matrix that satisfies the three metrics constrains: $|i_f|^2 = |j_f|^2, i_f \cdot i_f = 0$, and $i_1 = 1$. The motion matrix is defined as $M = [\ i_1 \ldots i_1 \mid j_1 \ldots j_F\ ]^T$.

Step 3: We recover the 3-D rotation angles θ_x, θ_y, and θ_z at every projection image from the motion matrix M. Entries of motion matrix M for each projection are defined to equal the first two rows of the arbitrary rotation matrix $R(\theta_z,\theta_y,\theta_x)$, where

$$\begin{aligned} R(\theta_z, \theta_y, \theta_x) &= R(\theta_z) \cdot R(\theta_y) \cdot R(\theta_x) \\ &= \begin{bmatrix} \cos\theta_Z & -\sin\theta_Z & 0 \\ \sin\theta_Z & \cos\theta_Z & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \cos\theta_Y & 0 & \sin\theta_Y \\ 0 & 1 & 0 \\ -\sin\theta_Y & 0 & \cos\theta_Y \end{bmatrix} \\ &\quad \times \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta_x & -\sin\theta_x \\ 0 & \sin\theta_x & \cos\theta_x \end{bmatrix} \end{aligned}$$

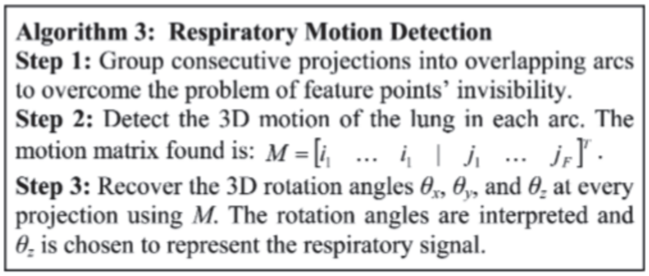
where $R(\theta_z)$, $R(\theta_y)$, and $R(\theta_x)$ correspond to the rotation matrices about the Z-axis, Y-axis, and X-axis, respectively, (So, $[\, i_f \ \ j_f \,]^T$) as shown at the bottom of the next page.

The rotation angle θ_y is interpreted as the gantry (orbital) motion of the X-ray source around the patient. θ_x is interpreted as the rotation around the X-axis which is not our concern in this study. θ_z is interpreted as the respiration signal because the lung anatomy is seen by the X-ray source as rotating around the Z-axis, as shown in Fig. 5. Algorithm 3 summarizes the steps used for respiratory motion detection.

D. *Projections Phase Sorting*

To sort the projections into phases, we use the respiratory signal θ_z as the input of the phase sorting method. The following steps show the process of breath phase sorting.

Step 1: The respiratory signal θ_z extracted is smoothed before phase sorting. Savitzky–Golay smoothing filter [36] is used to clear out the noisy respiratory signal. The technique uses a set of weighting coefficients $(w_{-m}, w_{-m-1}, \ldots, w_{m-1}, w_m)$ to carry out the smoothing operation. The use of these weights



Algorithm 3: Respiratory Motion Detection

Step 1: Group consecutive projections into overlapping arcs to overcome the problem of feature points' invisibility.

Step 2: Detect the 3D motion of the lung in each arc. The motion matrix found is: $M = [i_1 \ \ldots \ i_1 \mid j_1 \ \ldots \ j_F]^T$.

Step 3: Recover the 3D rotation angles θ_x, θ_y, and θ_z at every projection using M. The rotation angles are interpreted and θ_z is chosen to represent the respiratory signal.

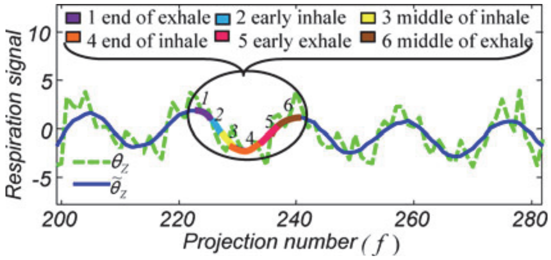


Fig. 6. Noisy (θ_z), smoothed respiratory signal ($\tilde{\theta}_z$), and phase sorting into six bins using the smoothed respiratory motion. The six colors in the oval correspond to each bin as shown on top of the figure.

C is equivalent to fitting the data to a polynomial. Thus, the smoothed data point $\tilde{\theta}_{z,f}$ at frame f is

$$\tilde{\theta}_{Z,f} = \sum_{i=-m}^{m} w_i \theta_{Z,f+i} \Big/ \sum_{i=-m}^{m} w_i. \tag{7}$$

Fig. 6 shows the respiratory signal smoothing and breath phase interpretation in one breath cycle.

Step 2: The smoothed respiratory signal $\tilde{\theta}_z$ is used in recovering breath phases. Let $H(f)$ denote the breath phase. We sort projections according to their breath phase by setting all peak projections to a phase of "1": $H(f) = 1$. The remaining projections between the two peaks are assigned to the other number of phase bins wanted by taking the total number of projections divided by the number of phases, and then sort the projections so that each bin contains roughly the same number of projections. Fig. 7 shows the breath phases extraction from the smoothed respiratory signal $\tilde{\theta}_Z$ in multiple breath cycles.

Step 3: We measure the breath phase sorting quality by comparing the sorted projections using LIFT and the ground truth signal. The phase shift, defined as $\Delta H(f)$, is measured between every corresponding phase "1" in both signals. The average phase shift is estimated for each arc and is defined as

$$\bar{H} = \sum_{c=1}^{n_c} \Delta H(f)/n_c \tag{8}$$

where n_c is the number of cycles. The standard deviation of phase shift is defined as $\sqrt{\sum_{c=1}^{n_c} (H(f) - \bar{H})^2 / n_c}$. Also, the breathing amplitude error is computed by dividing the average number of phase-shifted projections by the average number of projections in each cycle as follows:

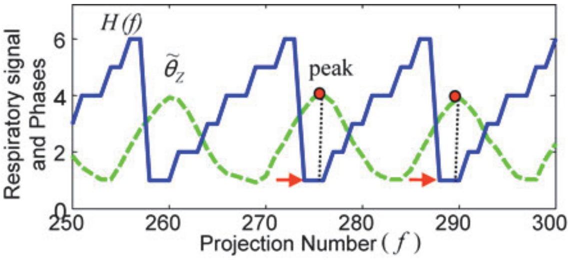


Fig. 7. Breath phase sorting of the projections into six bins based on the respiratory signal in multiple respiratory cycles. Projections are sorted into phases based on the respiration signal $\tilde{\theta}_z$ (dashed green). Phases $H(f)$ (solid blue) are interpreted as: 1 as the end of exhale, 2 as early inhale, 3 as middle of inhale, 4 as the end of inhale, 5 as early exhale, and 6 as middle of exhale.

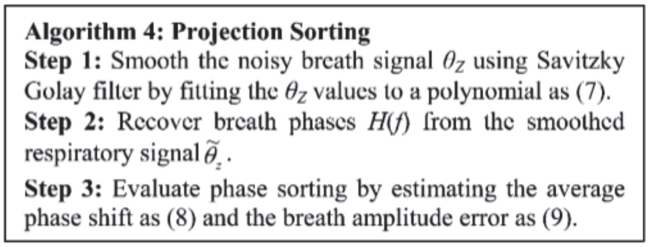
$$\bar{H} / (F/n_c). \tag{9}$$

III. EXPERIMENTAL RESULTS

Here, the experimental results of the proposed LIFT method are presented and analyzed. Section III-A shows datasets' specifications. Section III-B presents feature tracking and trajectory selection. Section III-C discusses respiratory motion detection and phase sorting. In Section III-D, the improved reliability and applicability is discussed.

A. Dataset Specifications

Four datasets were used to validate LIFT. The first dataset used has been taken under the following characteristics. The imaging system used consists of a radiation source and a detector panel which orbit in the *XZ* plane around the fixed point in space (which we will place at the world coordinate system origin). The actual distance of the source to origin is always fixed 1000 mm, and the virtual distance is also 1000 mm in this setup. The detector center to origin is also fixed 536 mm. The system is calibrated to provide a virtual image of known size at the origin by calibrating the physical pixel size. With this calibration, the virtual panel dimensions are 265.2 mm × 265.2 mm, and the pixel size is 0.518 mm/pixel. Patient 2–4s' datasets are similar, but acquired on a different vendor's imaging system. The



Algorithm 4: Projection Sorting

Step 1: Smooth the noisy breath signal θ_Z using Savitzky Golay filter by fitting the θ_Z values to a polynomial as (7).

Step 2: Recover breath phases $H(f)$ from the smoothed respiratory signal $\tilde{\theta}_z$.

Step 3: Evaluate phase sorting by estimating the average phase shift as (8) and the breath amplitude error as (9).

$$\text{So, } \begin{bmatrix} i_f & j_f \end{bmatrix}^T = R_2(\theta_z, \theta_y, \theta_x) = \begin{bmatrix} \cos\theta_Z \cos\theta_Y & \cos\theta_Z \sin\theta_Y \sin\theta_x - \sin\theta_Z \cos\theta_x & \cos\theta_Z \sin\theta_Y \cos\theta_x + \sin\theta_Z \sin\theta_x \\ \sin\theta_Z \cos\theta_Y & \sin\theta_Z \sin\theta_Y \sin\theta_x + \cos\theta_Z \cos\theta_x & \sin\theta_Z \sin\theta_Y \cos\theta_x - \cos\theta_Z \sin\theta_x \end{bmatrix}.$$

TABLE III
DATASET SPECIFICATIONS FOR MULTIPLE PATIENT DATASETS

Patient number	Number of projections	Projection size (pixel)	Pixel size (mm/pixel)	Source-Origin distance (mm)
1	701	512 × 512	0.518	1000
2	2396	768 × 1024	0.258	1000
3	2436	768 × 1024	0.258	1000
4	2300	768 × 1024	0.258	1000

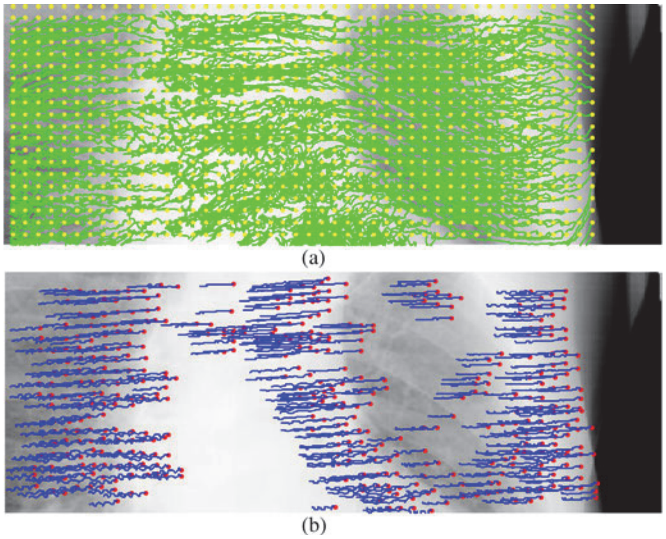


Fig. 8. Feature points selection and tracking: (a) Feature points detected in the first projection image of Arc 4 and trajectories through the arc. Yellow dots show the extracted pixels and green lines show the trajectories. (b) Selected trajectories that represent breathing in blue.

geometry is similar, but the virtual panel dimensions are 198.5 mm × 264.7 mm, and the pixel size is 0.258 mm/pixel. Patient 4's dataset showed some irregular breathing patterns, while patient 1–3's datasets have generally regular breathing patterns. The ground truth used for the datasets for patient 1 was the result of the diaphragm position-based method, while the ground truth for patient 2–4s' datasets was the internal markers trajectories. Four markers were used in patient 2's and patient 4's datasets and the averages of their trajectories through the projection images were estimated. One marker was used in patient 3's dataset. Table III shows the specifications of the datasets used in this study.

B. Feature Tracking and Trajectory Selection

Feature points were extracted in the first projection of the sequence on locations equally spaced by a specific number of pixels, 20 pixels in this experiment. Fig. 8 shows feature points tracking in Arc 4 of patient 1's dataset (between projection #250 and #500). The displacement shown is represented in pixels. In (a), yellow dots show the extracted feature points and green lines show the trajectories of those tracked points through the arc projections. The shape of the trajectories determines the motion of the area in which those trajectories reside. For example, trajectories residing in thoracic tissue regions tend to have a shape similar to a respiration curve, while trajectories residing in bony areas look like an orbital trajectory or unorganized shape. In (b), selected trajectories of the tracked points are shown.

TABLE IV
AVERAGE COMPACTNESS AND ISOLATION OF THE CHOSEN CLUSTER OF TRAJECTORIES IN MULTIPLE PATIENT DATASETS

Patient #	AVG Compactness (3)	AVG Isolation (4)
1	88.01 ± 10.19	43.20 ± 9.27
2	86.24 ± 12.53	47.53 ± 10.96
3	90.15 ± 9.84	45.14 ± 7.67
4	85.14 ± 9.54	42.52 ± 11.86

The average optical flow displacement was 0.51 pixels in patient 1's dataset, 0.37 pixels in patient 2's dataset, 0.23 pixels in patient 3's dataset, and 49.5 in patient 4's dataset. Our evaluation of optical flow performance has been limited to a qualitative judgment as [37] due to the lack of true displacement vectors. The displacement vectors estimated for patient 1's datasets in arc 4 have an average horizontal displacement larger than the vertical one. Those values were compared to Fig. 8 in which it appeared that the horizontal motion was larger than the vertical motion. Also, the results of the 3-D motion recovery, shown in Section III-C, proved that the optical flow algorithm used performed well and produced good results.

As previously mentioned, trajectories showing cyclical superior–inferior motion representing respiratory motion were selected according to the criteria followed in (2). Table IV presents the compactness and isolation measures [see (3) and (4)] applied to the resulting clusters in four patient datasets. As shown, the average compactness of the selected trajectories in the four patient datasets was around 87%, and the average isolation was around 44%.

C. Respiratory Motion Detection and Phase Sorting

To model the 3-D motion of the lung, structure from motion is used as described in Section II-C. Fig. 9 shows the extracted respiratory signal in selected arcs. The respiratory signal is the rotational angle about the Z-axis measured in degrees. LIFT-based respiratory signal was compared to the diaphragm position-based signal in (a) and to the marker-based signal in (b)–(d). The X-axis of the figure corresponds to the projection number, and the Y-axis corresponds to the degree of the rotation at each projection.

As shown in Fig. 9, LIFT respiratory motion matches the ground truth. To compare the accuracy of LIFT signal, we care about how much horizontal shift exists between the peaks of the signals extracted using LIFT and the ground truth. Smaller shifts imply stronger correlation between the two signals. The curves are scaled to have similar heights for comparison reasons. The difference in the height or the location of the curves of the different methods occurs due to the different surrogates used to represent the respiratory motion (e.g., diaphragm position versus rotational angle around the Z-axis). This difference has no importance as this information is not used to phase sort the projections.

Patient 1-3's datasets generally represented regular breathing. We anticipate that LIFT will be able to recover, or at least identify, irregular breathing patterns, as it is able to recover the actual 3-D motion through the sequence in any form, given accurate 2-D trajectories. Similar methods have been used to

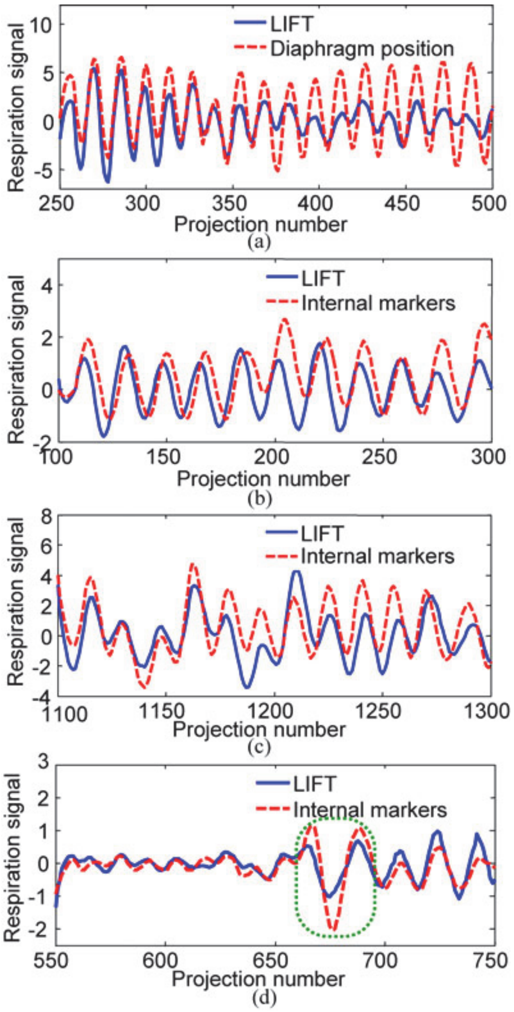


Fig. 9. Modeled respiratory motion in the datasets of (a) patient 1, (b) patient 2, (c) patient 3, and (d) patient 4. Patient 4 had changes in period and amplitude of breathing in this portion of the breathing trace, and the green rectangle shows a particularly large change in amplitude. The respiratory motion is the modeled 3-D rotation around the *Z*-axis measured in degrees. Peaks in LIFT respiratory signal and other methods are compared.

recover free-form 3-D motion [24]. In patient 4's dataset, an irregular portion of the breathing trace was detected using LIFT, as shown in Fig. 9(d). A deep breath around projection 680 was detected. This breath cycle is surrounded by a green rounded rectangle.

Fig. 10 shows respiratory phase sorting in the four patients' datasets as described in Section II-D. As shown, sorted projections using LIFT signal match the sorted projections using the ground truth signal especially in (a), (c), and (d). The average phase shift using LIFT and ground truth was quantitatively estimated as described in Section II-D.

Table V shows the quantitative accuracy of the breath phase sorting of the respiratory signal extracted of patient 1 using LIFT compared to the diaphragm position-based signal. We used the average and standard deviation of phase shift as described in (8). The phase shift between LIFT-based signal and the diaphragm-based signal was big in the first arc compared to the other arcs. This was because projection images in this arc have high contrast and the respiration signal extracted was less accurate compared to the other arcs.

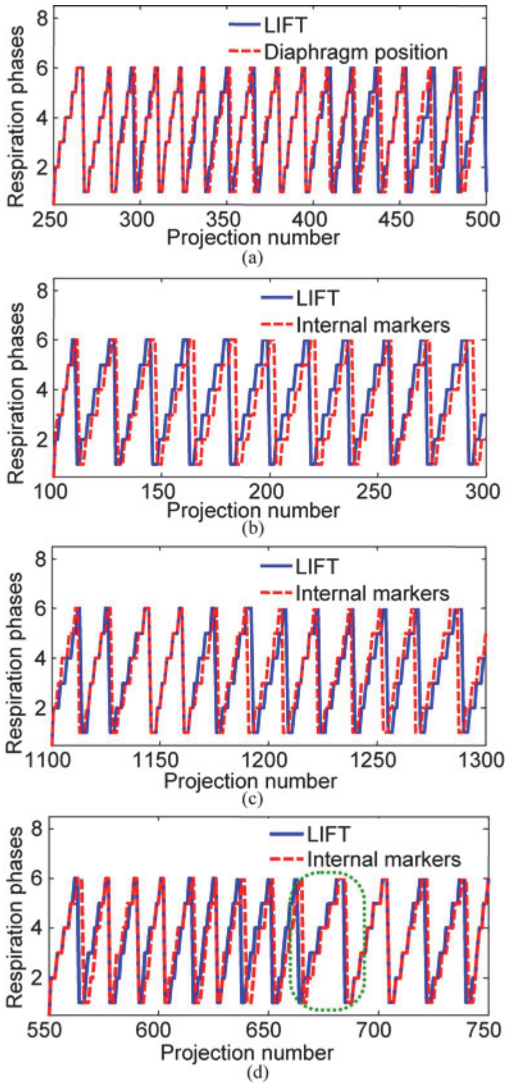


Fig. 10. Extracted respiration phases when the number of bins is 6 in (a) patient 1, (b) patient 2, (c) patient 3, and (d) patient 4. The green rounded rectangle in (d) corresponds to a cycle with large change in breathing amplitude. The X-axis is the projection number and the Y-axis is respiration phases ranging from 1 to 6. Respiration phases match when the respiration signal matches.

TABLE V
ERROR IN BREATH PHASE SORTING FOR PATIENT 1

Arc #	Frame#	AVG ± STD phase shift (8)
1	1-150	2.60 ± 0.84
2	100-300	1.29 ± 0.76
3	250-500	1.00 ± 0.82
4	450-701	1.86 ± 1.92

Table VI shows the quantitative measurement of phase shift between LIFT respiratory signal and the ground truth signal in all arcs for the four datasets. The criteria used for error measurement are discussed in Section II-D. We used the average and standard deviation of phase shift as in (8) and the average breathing amplitude error as in (9). As shown, the average phase shift measured in respiratory motion was around 1.68

TABLE VI
AVERAGE ERROR IN BREATH PHASE SORTING IN MULTIPLE DATASETS

Patient #	AVG phase shift (8)	AVG Breathing amplitude error (9)
1	1.68 ± 1.09	11.20 ± 7.27 %
2	3.04 ± 1.52	15.96 ± 7.98 %
3	1.25 ± 0.83	7.19 ± 4.77 %
4	1.14 ± 1.05	8.35 ± 7.96 %

compared to the diaphragm-based signal estimated in patient 1's dataset. Comparing to the internal marker-based signal, the average phase shift was 3.04 projections in patient 2, 1.25 in patient 3, and 1.14 in patient 4's datasets. As the average breathing amplitude error was estimated using the average phase shift, patient 3's case had the smallest error of 7.19% as an average, while it is 15.96% in patient 2's case. This shows that the computed signal using LIFT correlates to the diaphragm position-based signal in patient 1's case and to the internal marker's signal in patient 3's and patient 4's cases more than patient 2's case, if LIFT is performed on the entire projection. One of the reasons behind this is that the respiration signals extracted using markers may depend on the location of the markers. The respiration signal derived using LIFT is the 3-D rotation, about the Z-axis, of the lung tissues appearing in the raw CBCT projections. The locations of the feature points used in extracting the respiration signal affect the respiration signal extracted. Feature points in the lower part of the lung close to the diaphragm showed a strong oscillating motion more than the feature points in the upper part or edges of the lung. As LIFT finds a global 3-D motion of the flow intensity in the entire sequence, the effect of the stronger motion in the lower part of the lung dominates.

In patient 2's dataset, markers exist in the middle of the lung close to the bronchioles. The average phase shift between LIFT respiratory signal and markers' signal was 3.04 ± 1.52 projections. When we extracted the signal from a region of interest (ROI) surrounding the markers position, the phase shift was 1.6±1.9 projections which is less than the phase shift when extracting the respiratory motion from the entire projection. These results demonstrate that there were phase differences between different parts of the lung, and a single respiratory signal may not be optimal to completely characterize breathing motion. One advantage of LIFT is that it can be applied to a custom ROI surrounding the part of the lung containing the tumor, allowing more accurate characterization of the respiratory signal for this region.

Also, we applied LIFT on the inferior portion of the CBCT images that include the diaphragm. The average phase shift between the respiration signal using LIFT and the diaphragm position-based method was 1.1 ± 0.57 projections in arc 1 of patient 1's dataset, which was less than the phase shift excluding the diaphragm area (2.60 ± 0.84 projections). This result demonstrates phase shifts in the anatomy at different superior/inferior locations in the lung, which implies breathing phase obtained at the diaphragm may not be appropriate to represent phase at other locations.

Fig. 11 shows the reconstructed 4DCBCT images for patient 2 using the FDK algorithm [38]. We used projections that were sorted in the end of inhalation phase using the respiratory signal extracted by the implanted markers in (a1) and(a2), and LIFT in (b1) and (b2). The difference of the axial images (a1) and (b1) is shown in (c1) and difference of the coronal images (a2) and (b2) is shown in (c2). As shown in (c1) and (c2), most of the differences were not anatomical, but rather due to differences in the streaking (view-aliasing) artifact due to too few projections. Since each sorting algorithm selects slightly different projections for reconstruction, the streaks appear in different locations in the two images.

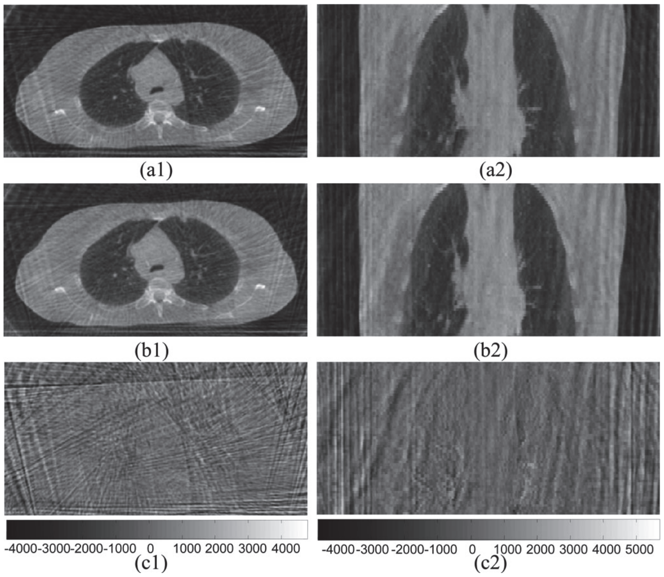

(a1) (a2)
(b1) (b2)
(c1) (c2)

Fig. 11. Reconstructed 4DCBCT images of patient 2, sorted in the end of inhalation phase using respiratory signal extracted using (a1) and (a2) implanted markers and (b1) and (b2) LIFT. (c1) Difference of the reconstructed images between the axial images (a1) and (b1). (c2) Difference of the reconstructed images between the coronal images (a2) and (b2).

TABLE VII
ACQUIRED DATA RATIO IN LIFT COMPARED TO OTHER EXISTING METHODS

Method	Samples	Acquired data ratio (%)
Diaphragm Position [8],[11]	Patient 1	100
Internal Markers [22], [23]	Patient 2: Marker 1, 2	100
	Marker 3	72.9
	Marker 4	70.0
	Patient 3: Marker 1	49.9
	Patient 4: Marker 1, 2	100
	Marker 3	68.8
	Marker 4	0
Proposed LIFT	Patients 1-4	100

D. Improved Reliability and Applicability

Here, we discuss the reliability and applicability of LIFT comparing to the other standard methods. Table VII shows the acquired data ratio using the diaphragm-based method [8], [11], the implanted markers method [22], [23], and LIFT. The acquired data ratio is the number of projections of the dataset from which the method was able to extract the respiratory signal, divided by the total of number of projections.

As shown in Table VII, the diaphragm-based method applied to patient 1's dataset was able to extract the respiratory signal in all projections. For the internal markers method, some of markers were not visible in some of the dataset projections. In patient 3's case, only one marker was used and was visible in around 50% of the projections, which prevented the respiration signal from being extracted in the rest 50% of the dataset projections. When having multiple markers implanted, visible markers' traces can be used as replacements to the invisible ones such as in the cases of patients 2 and 4. As LIFT extracts the respiratory motion from the organ tissues of the lung in any CBCT projection, it was able to generate a respiratory signal in 100% of the projection sets of the four patients. These results showed that LIFT was more reliable than the internal marker's method in terms of acquiring the respiration signal from projections.

The average time for running the entire LIFT procedure is around 26 min on a set of 1000 projections (of approximately 550×650 pixels as the analysis region) using a PC of Intel Core 2 Duo 2.4-GHz CPU and 2 GB of RAM. Running speed can be improved if LIFT is implemented on GPU with C++ platform or on a multiprocessor computer using MATLAB parallel processing. For implementation purposes, LIFT follows feature tracking and 3-D reconstruction approach which is well known and easy to implement. There are many freely available, optimized implementations that a developer can start from, such as OpenCV for optical flow computation [39]. For 3-D motion reconstruction, we used Tomasi and Kanade method [35] which is a mathematical approach consisting of well-described series of linear algebra computations. Agglomerative hierarchical clustering is described in Mathworks [40]. Thus, LIFT can be implemented and applied in the clinic for respiratory signal extraction from CBCT projections.

IV. CONCLUSION

We proposed a novel method for respiratory motion extraction and breath phase sorting using CBCT projections. Feature points were extracted and tracked to form point trajectories. Trajectories with shapes similar to breathing curve were selected to be used in the 3-D motion modeling module to recover the 3-D motion of the lung. The 3-D rotation around the Z-axis of the patient represented the respiratory motion in this study and the CBCT projections were then sorted according to the respiration signal.

Experimental results were conducted on datasets exhibiting regular breathing and some irregular breathing patterns. The respiratory motion extracted using LIFT was compared to the ones extracted using other standard methods. An average phase shift of 1.78 projections was estimated between LIFT-based signal and marker-based signal, and of 1.68 projections between LIFT-based signal and the diaphragm-based signal. The average breathing amplitude error of LIFT compared to the diaphragm-based method was 11.2%, while it is 10.68% compared to the internal markers method. LIFT was able to extract the respiration signal in all projections of all datasets without the dependence on a particular anatomical structure (such as the diaphragm).

ACKNOWLEDGMENT

The authors would like to thank J. Williamson.

REFERENCES

[1] A. Garde, L. Sörnmo, R. Jané, and B. F. Giraldo, "Correntropy-based spectral characterization of respiratory patterns in patients with Chronic Heart Failure," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 8, pp. 1964–1972, Aug. 2010.

[2] C. M. Ionescu, I. Muntean, J. A. Tenreiro-Machado, R. De Keyser, and M. Abrudean, "A theoretical study on modeling the respiratory tract with ladder networks by means of intrinsic fractal geometry," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 2, pp. 246–253, Feb. 2010.

[3] A. Yadollahi and Z. M. K. Moussavi, "A robust method for estimating respiratory flow using tracheal sounds entropy," *IEEE Trans. Biomed. Eng.*, vol. 53, no. 4, pp. 662–668, Apr. 2006.

[4] T. Li, L. Xing, P. Munro, C. McGuinness, M. Chao, Y. Yang, B. Loo, and A. Koong, "Four-dimensional cone-beam computed tomography using an on-board imager," *Med. Phys.*, vol. 33, no. 10, pp. 3825–3833, Sep. 2006.

[5] M. Y. Wang, C. R. Maurer, Jr., J. M. Fitzpatrick, and R. J. Maciunas, "An automatic technique for finding and localizing externally attached markers in CT and MR volume images of the head," *IEEE Trans. Biomed. Eng.*, vol. 43, no. 6, pp. 627–637, Jun. 1996.

[6] L. Pan, J. L. Prince, J. A. C. Lima, and N. F. Osman, "Fast tracking of cardiac motion using 3D-HARP," *IEEE Trans. Biomed. Eng.*, vol. 52, no. 8, pp. 1425–1435, Aug. 2005.

[7] A. S. Naini, T. Y. Lee, R. V. Patel, and A. Samani, "Estimation of lung's air volume and its variations throughout respiratory CT image sequences," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 1, pp. 152–158, Jan. 2011.

[8] J. Sonke, L. Zijp, P. Remeijer, and M. van Herk, "Respiratory correlated cone beam CT," *Med. Phys.*, vol. 32, no. 4, pp. 1176–1186, 2005.

[9] L. Dietrich, S. Jetter, T. Tucking, S. Nill, and U. Oelfke, "Linac-integrated 4D cone beam CT: First experimental results," *Phys. Med. Biol.*, vol. 51, pp. 2939–2952, May 2006.

[10] S. S. Vedam, V. R. Kini, P. J. Keall, V. Ramakrishnan, H. Mostafavi, and R. Mohan, "Quantifying the predictability of diaphragm motion during respiration with a noninvasive external marker," *Med. Phys.*, vol. 30, no. 4, pp. 505–513, Mar. 2003.

[11] S. Rit, D. Sarrut, and C. Ginestet, "Respiratory signal extraction for 4D CT imaging of the thorax from conebeam CT projections," *Med. Image Comput. Comput. Assist. Interv.*, vol. 8, no. Pt. 1, pp. 556–563, 2005.

[12] C. Ozhasoglu and M. J. Murphy, "Issues in respiratory motion compensation during external-beam radiotherapy," *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 52, no. 5, pp. 1389–1399, Apr. 2002.

[13] R. I. Berbeco, S. Nishioka, H. Shirato, G. T. Chen, and S. B. Jiang, "Residual motion of lung tumours in gated radiotherapy with external respiratory surrogates," *Phys. Med. Biol.*, vol. 50, no. 16, pp. 3655–3667, Aug. 2005.

[14] H. H. Liu, N. Koch, G. Starkschall, M. Jacobson, K. Forster, Z. Liao, R. Komaki, and C. W. Stevens, "Evaluation of internal lung motion for respiratory-gated radiotherapy using MRI: Part II-margin reduction of internal target volume," *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 60, no. 5, pp. 1473–1483, Dec. 2004.

[15] L. I. Cervino, A. K. Chao, A. Sandhu, and S. B. Jiang, "The diaphragm as an anatomic surrogate for lung tumor motion," *Phys. Med. Biol.*, vol. 54, no. 11, pp. 3529–3541, Jun. 2009.

[16] D. Gierga, G. Sharp, J. Brewer, M. Betke, C. Willett, and G. T. Chen, "Correlation between external and internal markers for abdominal tumors: implications for respiratory gating," *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 57, Suppl. 2, pp. S186–S187, Oct. 2003.

[17] J. D. Hoisak, K. E. Sixel, R. Tirona, P. C. Cheung, and J. P. Pignol, "Correlation of lung tumor motion with external surrogate indicators of respiration," *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 60, no. 4, pp. 1298–1306, Nov. 2004.

[18] D. Ionascu, S. B. Jiang, S. Nishioka, H. Shirato, and R. I. Berbeco, "Internal-external correlation investigations of respiratory induced motion of lung tumors," *Med. Phys.*, vol. 34, no. 10, pp. 3893–3903, Oct. 2007.

[19] S. Korreman, H. Mostafavi, Q. T. Le, and A. Boyer, "Comparison of respiratory surrogates for gated lung radiotherapy without internal fiducials," *Acta. Oncol.*, vol. 45, no. 7, pp. 935–942, Sep. 2006.

[20] Y. Tsunashima, T. Sakae, Y. Shioyama, K. Kagei, T. Terunuma, A. Nohtomi, and Y. Akine, "Correlation between the respiratory waveform measured using a respiratory sensor and 3D tumor motion in gated radiotherapy," *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 60, no. 3, pp. 951–958, Nov. 2004.

[21] H. Yan, F. F. Yin, G. P. Zhu, M. Ajlouni, and J. H. Kim, "The correlation evaluation of a tumor tracking system using multiple external markers," *Med. Phys.*, vol. 33, no. 11, pp. 4073–4084, Oct. 2006.

[22] T. E. Marchant, A. M. Amer, and C. J. Moore, "Measurement of inter and intra fraction organ motion in radiotherapy using cone beam CT projection images," *Phys. Med. Biol.*, vol. 53, pp. 1087–1098, Feb. 2008.

[23] P. R. Poulsen, B. Cho, and P. J. Keall, "A method to estimate mean position, motion magnitude, motion correlation, and trajectory of a tumor from cone-beam CT projections for image-guided radiotherapy," *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 72, pp. 1587–1596, Dec. 2008.

[24] B. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in *Proc. 7th Int. Joint Conf. Artif. Intell.*, Vancouver, BC, Canada, 1981, pp. 674–679.

[25] F. Bergner, T. Berkus, M. Oelhafen, P. Kunz, T. Pan, and M. Kachelrieß, "Autoadaptive phase-correlated (AAPC) reconstruction for 4D CBCT," *Med. Phys.*, vol. 36, no. 12, pp. 5695–5706, Dec. 2009.

[26] B. K. Horn and B. G. Schunk, "Determining optical flow," *Artif. Intell.*, vol. 17, pp. 185–203, Aug. 1981.

[27] C. Wachinger, M. Yigitsoy, E.-J. Rijkhorst, and N. Navab, "Manifold learning for image-based breathing gating in ultrasound and MRI," *Med. Image Anal.*, vol. 16, no. 4, pp. 806–818, May 2012.

[28] R. A. C. Siochi, "Deriving motion from megavoltage localization cone beam computed tomography scans," *Phys. Med. Biol.*, vol. 54, no. 13, pp. 4195–4212, Jun. 2009.

[29] A. Kavanagh, P. M. Evans, V. N. Hansen, and S. Webb, "Obtaining breathing patterns from any sequential thoracic x-ray image set," *Phys. Med. Biol.*, vol. 54, no. 16, pp. 4879–4888, Aug. 2009.

[30] I. Vergalasova, J. Cai, and F.-F. Yin, "A novel technique for markerless, self-sorted 4D-CBCT: Feasibility study," *Med. Phys.*, vol. 39, no. 3, pp. 1442–1451, Mar. 2012.

[31] C. Tomasi and T. Kanade, "Detection and tracking of point features," Carnegie Mellon Univ., Pittsburgh, PA, Tech. Rep., CMU-CS-91-132, Apr. 1991.

[32] J. Shi and C. Tomasi, "Good features to track," in *Proc. IEEE Comput. Soc. Conf. Comput. Vision Pattern Recognit.*, New York, 1994, pp. 593–600.

[33] R. R. Sokal and P. H. A. Sneath, *Principles of Numerical Taxonomy*. San Francisco, CA: Freeman, 1963.

[34] R. Dubes and A. K. Jain, "Validity studies in clustering methodologies," *Pattern Recognit.*, vol. 11, no. 4, pp. 235–254, 1979.

[35] C. Tomasi and T. Kanade, "Shape and motion from image streams under orthography: A factorization method," *Int. J. Comput. Vis.*, vol. 9, no. 2, pp. 137–154, Nov. 1992.

[36] A. Savitzky and M. J. E. Golay *Anal. Chem.*, vol. 36, pp. 1627–1639, 1964.

[37] J. Barron, D. Fleet, and S. Beauchemin, "Performance of optical flow techniques," *Int. J. Comp. Vis.*, vol. 12, no. 1, pp. 43–77, 1994.

[38] L. A. Feldkamp, L. C. Davis, and J. W. Kress, "Practical cone-beam algorithm," *J. Opt. Soc. Amer. A-Opt. Image Sci. Vis.*, vol. 1, pp. 612–619, 1984.

[39] J.-Y. Bouguet, "Pyramidal implementation of the Lucas Kanade feature tracker," OpenCv Documentation, Microprocessor Research Labs, Intel Corp., 2000.

[40] MATLAB: Online help of hierarchal clustering-Release R2012a (2012). [Online]. Available: http://www.mathworks.com

Salam Dhou (S'11) received the B.S and M.S degrees in computer science from Jordan University of Science and Technology (JUST), Irbid, Jordan, in 2004 and 2007, respectively. She is currently working toward the Ph.D. degree in electrical and computer engineering from Virginia Commonwealth University, Richmond.

From 2007 to 2008, she was an instructor in the Department of Computer Science, JUST. Her research interests include data mining, target tracking, and medical imaging.

Yuichi Motai (M'01) received the B.Eng. degree in instrumentation engineering from Keio University, Tokyo, Japan, in 1991, the M.Eng. degree in applied systems science from Kyoto University, Kyoto, Japan, in 1993, and the Ph.D. degree in electrical and computer engineering from Purdue University, West Lafayette, IN, in 2002.

He is currently an Assistant Professor of electrical and computer engineering at Virginia Commonwealth University, Richmond. His research interests include the broad area of sensory intelligence; particularly in medical imaging, pattern recognition, computer vision, and sensory-based robotics.

Geoffrey D. Hugo received the B.S. degree in physics from Pennsylvania State University, University Park, in 1999, and the M.S. and Ph.D. degrees in biomedical physics from the University of California, Los Angeles, in 2002 and 2003, respectively.

He is board certified in radiation therapy physics and is currently an Assistant Professor of radiation oncology at Virginia Commonwealth University, Richmond, VA. His research interests include image-guided and adaptive radiation therapy, motion management, and deformable image registration.