



HOKKAIDO UNIVERSITY

Title	Markerless tracking of tumor and tissues: A motion model approach
Author(s)	Cheung, Ling Fung; Fujitaka, Shinichirou; Fujii, Takaaki et al.
Citation	Medical physics, 52(2), 1193-1206 https://doi.org/10.1002/mp.17524
Issue Date	2025-02
Doc URL	https://hdl.handle.net/2115/96194
Rights	This is the peer reviewed version of the following article: Cheung LF, Fujitaka S, Fujii T, Miyamoto N, Takao S. Markerless tracking of tumor and tissues: A motion model approach. Med Phys. 2025; 52: 1193-1206, which has been published in final form at https://doi.org/10.1002/mp.17524 . This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Use of Self-Archived Versions. This article may not be enhanced, enriched or otherwise transformed into a derivative work, without express permission from Wiley or by statutory rights under applicable legislation. Copyright notices must not be removed, obscured or modified. The article must be linked to Wiley's version of record on Wiley Online Library and any embedding, framing or otherwise making available the article or pages thereof by third parties from platforms, services and websites other than Wiley Online Library must be prohibited.
Type	journal article
File Information	MedPhys2024_r13_Cheung.pdf



Markerless Tracking of Tumor and Tissues: A Motion Model Approach

Ling Fung Cheung¹, Shinichirou Fujitaka¹, Takaaki Fujii¹, Naoki Miyamoto^{2,3} and Seishin Takao^{2,3}

¹Electromagnetic Application Systems Research Department, Research and Development Group, Hitachi Ltd., Hitachi, Ibaraki, Japan

²Division of Quantum Science and Engineering, Faculty of Engineering, Hokkaido University, Sapporo, Hokkaido, Japan

³Department of Medical Physics, Hokkaido University Hospital, Sapporo, Hokkaido, Japan

ABSTRACT

Background: Respiratory motion management is essential in order to achieve high-precision radiotherapy. Markerless motion tracking of tumor can provide a non-invasive way to manage respiratory motion, thereby enhancing the treatment accuracy. However, the low contrast in real-time X-ray images for image guidance limits the application of markerless tracking.

Purpose: We present a novel approach based on a motion model to perform markerless tracking of tumor and surrounding tissues even when they have low contrast in real-time X-ray images.

Methods: A proof-of-concept validation of the method has been performed using digital and physical phantoms at breathing conditions that are significantly different than the planning stage. A motion model is first constructed by performing principal component analysis (PCA) on the planning 4DCT. During treatment, the motion of a surrogate is tracked and used as the input of the motion model, which generates a 3D real-time volume estimation. Such 3D estimation is then projected to 2D to create digitally reconstructed radiographs (DRRs). The relationships between the real-time DRRs, reference DRRs, and reference X-ray images are first established to simulate 2D real-

time images from the real-time volume. The registration between the simulated 2D real-time images and real-time X-ray images corrects the initial motion model estimation to ensure the estimated volume match the real-time condition.

Results: In digital phantom, the Dice index of pancreas was improved from 0.74 to 0.78 after correction using real-time DRRs in fully inhaled phase. Validation on lung and pancreas are performed in physical phantom with two motion traces. The surrogate-tumor relationships were intentionally altered to generate large target localization errors due to the differences in body condition between treatment planning stage and during treatment. The real-time correction for the estimated 3D real-time volume was performed using a pair of 2D X-ray images. For the deep breathing motion trace, the tumor localization mean absolute error (MAE) throughout the tracking decreases from around 3 mm to less than 1 mm after correction. For the shallow breathing motion trace with a 1.7 mm baseline shift, the tumor localization mean absolute error (MAE) throughout the tracking decreases from around 1.5 mm to less than 1 mm after correction.

Conclusions: The method combines the detailed structural information from planning 4DCT and real-time information from real-time X-ray images through a motion model. The matching between the real-time model estimation and 2D real-time images are performed in the same modality so that it can be applied to regions with low contrast in the images. The real-time images successfully corrected the initial motion model estimations in our proof-of-concept validation. This suggests the potential to perform markerless tracking in low-contrast region using a motion model.

Correspondence

Ling Fung Cheung,

- 50 Electromagnetic Application Systems Research Department, Research and Development Group, Hitachi
51 Ltd., Hitachi, Ibaraki, Japan
52 Email: lingfung.cheung.wg@hitachi.com

1 INTRODUCTION

Respiratory motion is one of the major challenges in achieving high-precision radiotherapy. For particle beam therapy (PBT), because doses are more focused to a specific region comparing to conventional radiotherapy, motion management is especially important to ensure optimal dose delivery. External tracking systems have been widely used for motion monitoring. Through combination with gating, treatment beam can be turned on only at certain breathing phase, thus increasing the treatment accuracy. However, the correlations between the motions of external surrogate and tumor are not necessarily stable, which can lead to target localization error.¹

Image-guided techniques can significantly increase the accuracy of motion tracking. Currently, internal fiducial markers are typically used to perform motion tracking through real-time X-ray images.¹⁻³ However, there are potential factors that might deteriorate the motion tracking accuracy when using internal markers, such as marker migration or image artifacts created by the markers. More importantly, it requires invasive procedures to implant makers inside patient's body, which might introduce complications and risks to patients. Therefore, it is more desirable to perform motion tracking without the use of internal markers.

Numerous efforts have been made on markerless tracking research.⁴⁻¹⁵ Particularly, significant progress has been achieved on tracking lung targets.⁵⁻¹⁴ Many methods utilize template-matching between the digitally reconstructed radiograph (DRR) and real-time X-ray image to determine the 3D tumor position.⁸⁻¹⁴ Some have already been applied clinically and achieved good tracking accuracy.¹¹⁻¹⁴

While the organs in abdomen such as liver and pancreas also undergo vigorous motions during breathing, there has been limited progress regarding X-ray based markerless tracking of these organs. This is probably due to the low contrast of tumor and tissues in real-time X-ray images, which makes effective template matching very difficult due to the difference in modalities between DRRs and X-ray images.

Recently, a patient-specific deep learning approach has been proposed to perform markerless tracking for pancreatic tumor.¹⁵ While AI-based methods have the power to extract clues regarding 3D tumor position from the limited information in 2D in-beam images, AI-related issues such as lack of process transparency and strong dependency on training data still need to be addressed. There have been studies that combine the offline and online training data for tumor position prediction, showing promising results that might increase the robustness of AI-based methods.¹⁶

In addition to positional displacements, breathing can also cause tumor deformation. It has been reported that the changes in tumor shape during breathing can be as large as 5 mm.¹⁷ In fact, other surrounding anatomic changes such as swallowing, peristalsis, and bladder filling can also cause the deformation. For PBT, the motion and deformation of surrounding tissues and gas along the beam path can largely affect the actual beam range, which make it even more challenging to deliver accurate dose during treatment.^{3,18}

Motion model has been proposed as a method for motion tracking that allows the extraction of real-time 3D information.¹⁹⁻²⁴ It can correlate the real-time motion of a surrogate with that of tumor and other tissues by taking advantage of the pre-treatment information, such as 4DCT. Principle component analysis (PCA) can be used to create a motion model from the deformable image registrations (DIR) between the phases in

4DCT.²²⁻²⁶ Through tracking the surrogate motion during treatment, the motion model generates a 3D real-time volume, which allows the calculation of real-time beam range in PBT. Li et al. proposed a method to generate real-time projection images from the volume created by a PCA lung motion model and perform optimization to minimize the difference between the projection images and real-time images.^{23,24} However, like the template-matching based methods, the method still relies on matching the CT-based projection images and real-time X-ray images directly. Such difference in image modalities limits its applicable region to X-ray clear regions such as lung. Recently, motion model-based approaches have been developed for MRI-based tracking, which extends the information from 2D cine-MRI to 3D volume.^{26,27} While MRI systems provide good soft-tissue contrasts and radiation-free image-guided treatment, the in-beam images contain only 2D information. Therefore, it is difficult to ensure the accuracy of the estimated 3D real-time volume in the MRI-based methods.

We propose a novel framework to perform markerless tracking of tumor and surrounding tissues for X-ray low contrast region using a motion model approach. Our method aims at connecting the structural information from planning 4DCT and real-time 3D information from a pair of in-beam 2D images through a motion model and a real-time correction process. The real-time correction between the model estimation and in-beam images are designed to perform in the same modality, so that the method can be applied to regions with low contrast in X-ray.

2 MATERIALS AND METHODS

2.1 Motion model

Our motion model is created in similar ways as proposed by previous studies.²²⁻²⁶ Figure 1 shows the schematic illustration on the creation of motion model and its application for motion tracking using a surrogate motion as the input. Briefly, PCA is applied to the deformation vector fields (DVF) obtained by deformably registering the 4DCT image of the reference phase to the images of other breathing phases. The end-of-exhale phase (T50) is used as the reference phase because it is more stable and contains less motion artifacts than other phases in the 4DCT image. The selected surrogate motion $s(t_{Tn0})$ from the DRR images created from the 4DCT are used as the surrogate input for the PCA. To perform PCA, the DVF and surrogate input are first centered by subtracting their mean motion. In this proof-of-concept study, one principal component is used. The PCA yields a matrix B , which is the motion model that takes the centered surrogate motion $\tilde{s}(t)$ as input and generates a centered real-time deformation field $\widehat{DVF}(t)$.

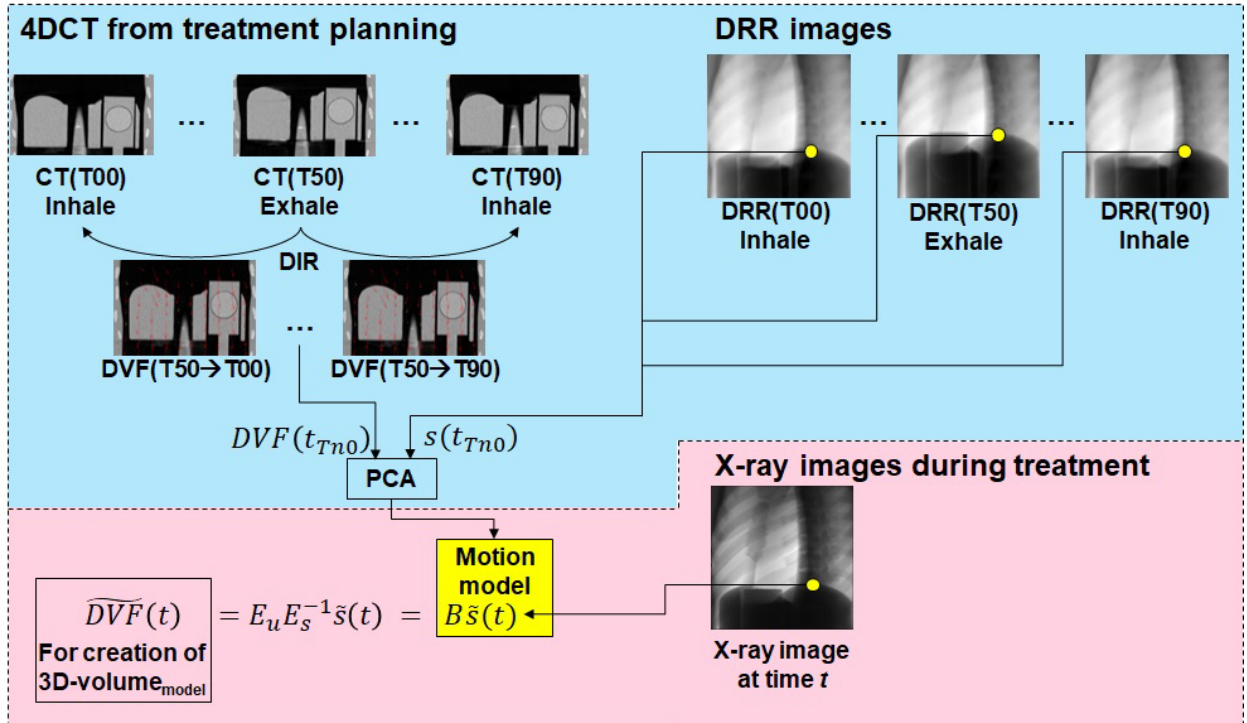


FIGURE 1 Schematic illustration on the creation of motion model and its application for motion tracking using a surrogate motion (yellow dot) as the input. $\tilde{s}(t)$ is the centered real-time surrogate motion. $\widetilde{DVF}(t)$ is the centered real-time deformation vector field.

The motion model relates the motion of a surrogate with that of tumor and other tissues. As a result, a real-time 3D-DVF $\widetilde{DVF}(t)$ can be estimated by monitoring the external or internal surrogate motions in real-time X-ray images during treatment. This 3D-DVF is then applied to deform the reference CT to generate a 3D real-time (RT) volume.

In general, the motion of an internal surrogate, such as diaphragm, correlates better with tumor motion than that of an external surrogate, such as body surface.²⁸ Therefore, in our approach, the top point of diaphragm is used as the surrogate. Using a pair of X-ray image detectors, its 3D position can be tracked in real-time during treatment.

2.2 Limitation of motion model for real-time tracking

By applying the motion model, the motions of tumor and tissues at any given breathing condition can be simulated. However, such motion simulations are limited by the 4DCT taken during the planning stage. While the inter-fractional changes can be addressed by pre-treatment 4DCBCT or other adaptive techniques, the limitations listed below still exist:

1. The temporal resolution of 4DCT limits the information obtained regarding tumor and tissue motions. As a result, the motions between each phase of the 4DCT cannot be fully described by the motion model. Likewise, if the motion exceeds the breathing amplitude when the 4DCT was taken, the motion model might generate large errors.

2. The motion model relies on the motion correlation established between the surrogate and tissues including tumor that appeared in the 4DCT. If such correlation alters between treatment fractions or during treatment, the error of the motion model estimation increases.

3. The instantaneous movements of the adjacent organs during treatment can cause unpredictable changes. For examples, the peristalsis of the digestive tract and the gas motions can affect the shapes and positions of a nearby tumor. These sudden changes can also largely alter the beam range if it affects the tissues along the beam path.

Due to these limitations, real-time information is necessary to correct the motion model estimation.

2.3 Real-time correction

Real-time X-ray imaging is a standard way to obtain real-time information regarding internal motion. Previous approaches achieve motion tracking through generation of real-time DRRs from the real-time 3D volume estimated by the motion model.^{23,24} These DRRs represent the 2D image estimation of the motion model in real-time. Through the matching between the real-time DRRs and real-time X-ray images, the original estimation can be corrected. However, DRR is synthetic image generated by projecting the CT volume onto 2D planes, which is different from the actual X-ray images detected from the flat panel detectors (FPD), due to the conversion of CT numbers to attenuation coefficients, noise, detector response, and other errors. Such image modality difference between DRRs and X-ray images makes it difficult to perform effective matching in regions other than lung. In

fact, even for lung tumor, it is still challenging to achieve markerless tracking successfully based on the matching between DRRs and X-ray images, especially for small lung tumor with low contrast.²⁹

Here we propose a procedure so that matching between the estimated images and real-time images can be performed in the same modality. The schematic workflow is presented in Figure 2.

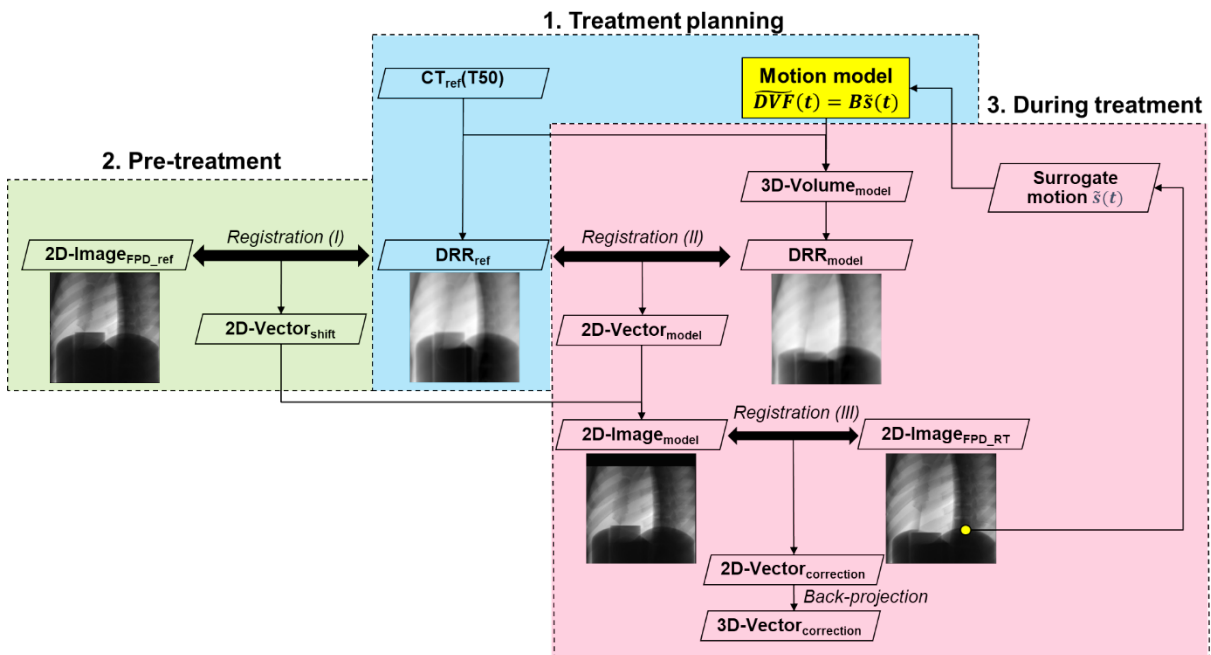


FIGURE 2 Schematic workflow to perform real-time correction of motion model estimation with the estimated images $2D-Image_{model}$ and real-time images $2D-Image_{FPD_RT}$ in the same modality. The yellow dot represents the surrogate.

The reference CT, $CT_{ref}(T50)$ taken at the end-of-exhale phase in treatment planning stage, is first projected to create reference DRR (DRR_{ref}). During patient positioning and before the treatment begins, the reference 2D X-ray images ($2D-Image_{FPD_ref}$) from the

FPD are taken at the end-of-exhale phase, which is the reference state. The patient-setup is performed using multi marker lines painted on patient's surface and the lasers in the treatment room for coarse alignment. Then the fine patient-alignment is performed automatically or manually using the bony structure information from the orthogonal X-ray images or cone-beam CT image.

Next, a first registration (I) is performed between the $2D\text{-Image}_{\text{FPD_ref}}$ and DRR_{ref} . Such registration establishes a connection between the two modalities, as well as the 2D-3D relationship, i.e. the correspondence between the 4DCT coordinate systems and the coordinate systems of the 2D X-ray images. Since the registration is done offline, it can be performed in a way so that a strong connection between the two modalities is established. During treatment, at the time when motion is being tracked, the surrogate motion is used as the input of the motion model. The motion model generates a 3D RT volume ($3D\text{-volume}_{\text{model}}$), which can then be projected to 2D to form a real-time DRR ($\text{DRR}_{\text{model}}$). A second registration (II) is performed between the DRR_{ref} and $\text{DRR}_{\text{model}}$ so that a 2D transformation vector describing the 2D motion estimation from motion model is obtained ($2D\text{-Vector}_{\text{model}}$). The $2D\text{-Vector}_{\text{model}}$ is then applied to the $2D\text{-Image}_{\text{FPD_ref}}$ to generate $2D\text{-Image}_{\text{model}}$. Since the $2D\text{-Image}_{\text{model}}$ has the same modality as the real-time $2D\text{-Image}_{\text{FPD_RT}}$, a better registration can be performed comparing to the conventional matching between DRR and X-ray images. Such third registration (III) generates a correction vector $2D\text{-Vector}_{\text{correction}}$, which can be back-projected according to the imager geometry to obtain a $3D\text{-Vector}_{\text{correction}}$. Such 3D correction vector corrects the motion model DVF, thus correcting the previously estimated 3D RT volume, yielding a corrected tumor position.

2.4 Proof-of-concept validation

Both digital and physical phantoms were used to validate the feasibility of the proposed method. To simplify the 2D-3D conversions, all the registrations in 2D were performed in rigid manner.

2.4.1 Digital phantom

A digital phantom, XCAT, was used to simulate 3D volumes in various breathing conditions.³⁰ Table 1 shows the major parameters set to simulate the breathing conditions for model creation and validation. Normal breathing condition with 12 mm and 20 mm breathing amplitudes in AP and SI directions were used to create the motion model. Ten phases of the 4DCT were generated and the reference phase (end-of-exhale) was deformably registered to the other nine phases using Plastimatch.³¹ The resulting DVFs were then used to create the motion model.

For validation purpose, deep breathing with 20 mm and 40 mm breathing amplitudes in AP and SI directions, respectively, were used to simulate condition which is not described by the 4DCT that constructed the motion model. Twenty phases of volume were created and defined as the ground truth. The volumes were projected to 2D at imaging angles of 180° and 270° to create the DRRs using a home-made program. In this digital simulation, the DRRs act as 2D real-time images to correct the original motion model estimation. The estimated volumes before and after the correction are then compared with the ground truths. Dice similarity coefficient (DSC) is used to measure the similarity

between the ground truth pancreas volume X and estimated pancreas volume Y with and without the correction using the formula below:

$$DSC = \frac{2|X \cap Y|}{|X| + |Y|}$$

TABLE 1 Major parameters set to simulate the breathing conditions for model creation and validation in digital phantom.

Process	Breathing condition	AP amplitude (mm)	SI amplitude (mm)	No. of phases
Model creation	Normal	12	20	10
Validation	Deep	20	40	20

2.4.2 Physical phantom

To perform validation in physical phantom, a dynamic thorax phantom made by Kyoto Kagaku was used. The experiment setup is shown in Figure 3(a). The original phantom was designed for the simulation of human thorax environment, with tumor models of various CT numbers and a low-density rod, which moves independently. In this study for lung, a 12 mm diameter tumor model with CT number of +100 HU is placed inside the rod. A mediastinum, as shown in Figure 3(b), was placed inside the phantom while the diaphragm was moving. A thinner diaphragm with its shape extracted from XCAT was 3D-printed, with a hole to allow the rod moving independently. Acrylic was selected as the 3D-printing material, with a CT number around +30 HU. The 3D-printed diaphragm and the

tumor rod set from Kyoto Kagaku for thorax and their corresponding positions inside the phantom are shown in Figure 3(c).

In addition to the simulation of thorax environment, we also tried to simulate the abdomen environment, which contains a high density of tissues. Similar to the 3D-printed diaphragm for lung, a thicker diaphragm with its shape extracted from XCAT was 3D-printed with acrylic material. Moreover, the rod was also 3D-printed to host a 50 mm diameter tumor model with CT number of +150 HU. The 3D-printed diaphragm and tumor rod for abdomen are shown in Figure 3(d).

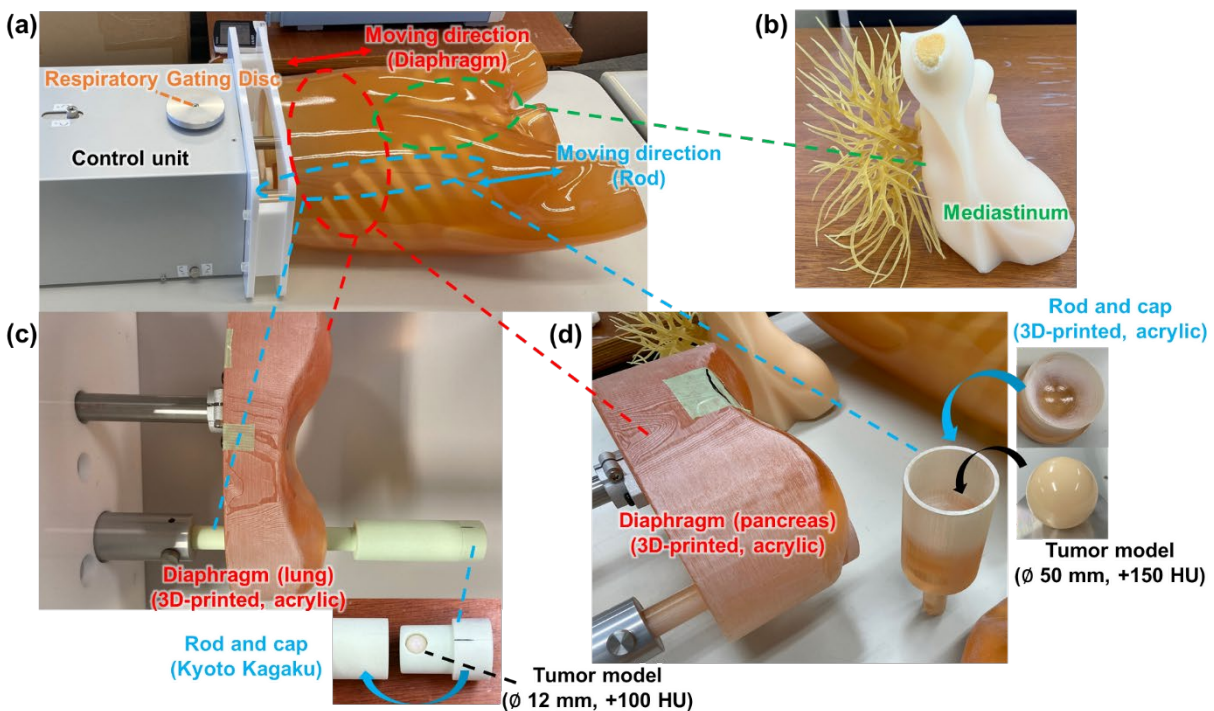


FIGURE 3 (a) Experiment setup of the physical phantom, (b) the mediastinum, and the diaphragm and rod designed for simulating (c) thorax environment with a lung tumor and (d) abdomen environment with a pancreas tumor.

Table 2 shows the breathing amplitudes of the two motion traces used in the validations. The motion model was first constructed using the 4DCT taken under normal breathing conditions with an amplitude of 17.9 mm for both diaphragm and tumor (Motion trace 0). The 4DCT was taken using SIEMENS SOMATOM confidence in Hokkaido University Hospital, with a marker block placed on the “Respiratory Gating Disc” tracked for phase binning. Due to the limitation of the motion control unit, the motion was limited to SI direction only. Two motion traces with amplitudes of 12.5 and 28.6 mm for both diaphragm and tumor were assessed to measure the inherent tracking error of diaphragm and tumor, which are shown in the Supplementary Information. Next, two other breathing patterns with significantly different breathing amplitudes were applied. In these two breathing patterns, the tumor rod and diaphragm were intentionally set to move differently to simulate cases which the correlation between the motions of surrogate and tumor has changed after the planning 4DCT was taken. For Motion trace 1, which aims at simulating a deep breathing condition, the moving amplitude for diaphragm was 27.6 mm and that for the tumor was 33.2 mm. While for Motion trace 2, which aims at simulating a shallow breathing condition with 1.7 mm baseline shift, the moving amplitude for diaphragm was 27.6 mm and that for the tumor was 33.2 mm. These motion traces are applied to the model creation and validation processes for both simulated lung and pancreas tumors.

TABLE 2 Major parameters set to simulate the breathing conditions for model creation and validation in physical phantom.

Process	Breathing condition	Diaphragm amplitude (mm)	Tumor amplitude (mm)
---------	---------------------	-----------------------------	-------------------------

Model creation (Motion trace 0)	Normal	17.9	17.9
Validation 1 (Motion trace 1)	Deep	27.6	33.2
Validation 2 (Motion trace 2)	Shallow (1.7 mm baseline shift)	12.5 + 1.7	15.3 + 1.7

The experiment was conducted in currently operating PBT treatment facility in Hokkaido University Hospital. The facility has the ability to perform gated PBT treatment based on the tracking result from internal fiducial marker.³ Therefore, after applying our new markerless tracking method to obtain tumor position, it is straightforward to apply to PBT treatment using the tracked position as the gating signal. Figure 4 shows the experimental setup for taking real-time X-ray images. A pair of X-ray tubes and FPD were used for imaging at a rate of 30 Hz to gather as much information as possible for validation. For lung simulation, the tube voltage and tube current were set to be 100 kV and 50 mA, respectively, with an exposure time of 3 ms. For pancreas simulation, the tube voltage and tube current were set to be 110 kV and 60 mA, respectively, with an exposure time of 3 ms. Such imaging conditions were typically used for motion tracking using internal markers in current treatment of lung and pancreatic tumors. In clinical practice, the imaging rate is typically 15 Hz. For typical clinical cases in Hokkaido University Hospital, the imaging dose would be less than 5% of the dose prescribed for the target, which is within the recommended threshold in the AAPM Task Group 180.³² The average contrast-to-noise ratio (CNR) of the two imagers for the lung tumor and pancreas tumor are 1.39

and 0.42, respectively. The details of the calculation of CNR are given in the Supplementary Information.

After specifying the motion of the phantom, X-ray images were taken synchronously at two imaging angles: 135° and 225° . The breathing motion repeats at a period of 5 seconds. We have recorded the image series for 40 seconds. Since the breathing patterns are constant, we only performed analysis on one breathing cycle between the 25 seconds to 30 seconds, in the middle of the image series taken from the imagers. In addition, for simplicity in this proof-of-concept study, we performed model correction every 500 ms among the analyzed period. DRRs are generated using the same geometry as the X-ray images taken. Since all motions are rigid, registration (I) described in Figure 2 was implemented in rigid manner and the correspondence is straightforward. Registration (II) was then carried out between DRR_{ref} and DRR_{model} to simulate the real-time X-ray images ($2D-Image_{model}$). Registration (III) between the ten $2D-Image_{FPD_RT}$ and $2D-Image_{model}$ generates a pair of 2D correction vectors, which are back-projected to 3D for the correction of the estimated 3D RT volume.

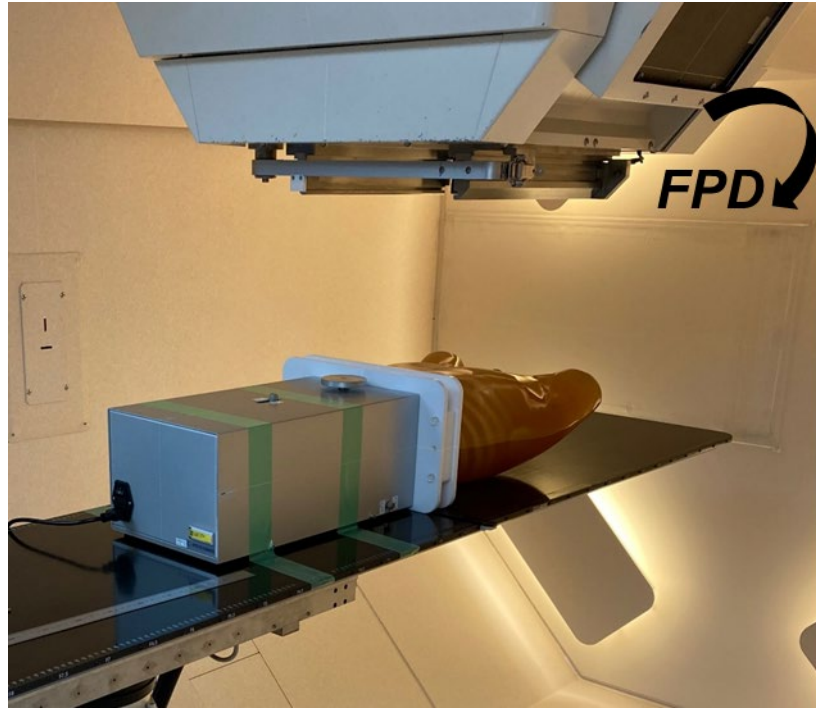


FIGURE 4 Experimental setup for data acquisition of X-ray images for motion tracking.

3 RESULTS

3.1 Digital phantom

We first validate the concept of the proposed method in digital phantom. Figure 6 shows the 4DCT images generated by XCAT. Sagittal plane was presented to visualize the breathing amplitudes at T50 phase (Figure 5a), T00 phase under the normal breathing (Figure 5b) and deep breathing (Figure 5c) motion as described in Table 1. In this study, the CT images for T50 phase are the same for both the normal breathing and deep breathing conditions.

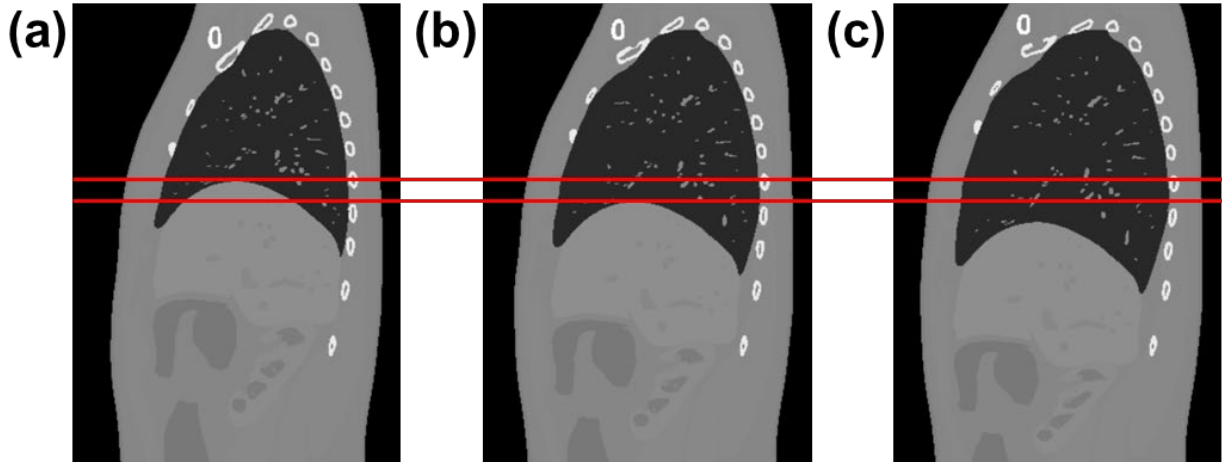


FIGURE 5 4DCT images generated by XCAT in the sagittal plane at (a) T50 phase, and T00 phase under the (b) normal breathing and (c) deep breathing motion. The red lines show the relative diaphragm positions.

Next, to create 3D real-time volume using a motion model, a surrogate input is needed. The diaphragm top point in DRR was tracked and used as the surrogate input of the motion model. The 3D real-time volume generated by the motion model was then projected to 2D plane with a specified angle. The DRR created by projecting the ground truth volume of the deep breathing T00 phase (Figure 6a) is compared with the DRR created by projecting the model volume at an imaging angle 270° (Figure 6b). The DRR_{model} shows signs of distortion near the diaphragm edges. This is due to the quality of DIR and excessive extrapolation of the motion model estimation, which was created under the condition with halved breathing amplitude.

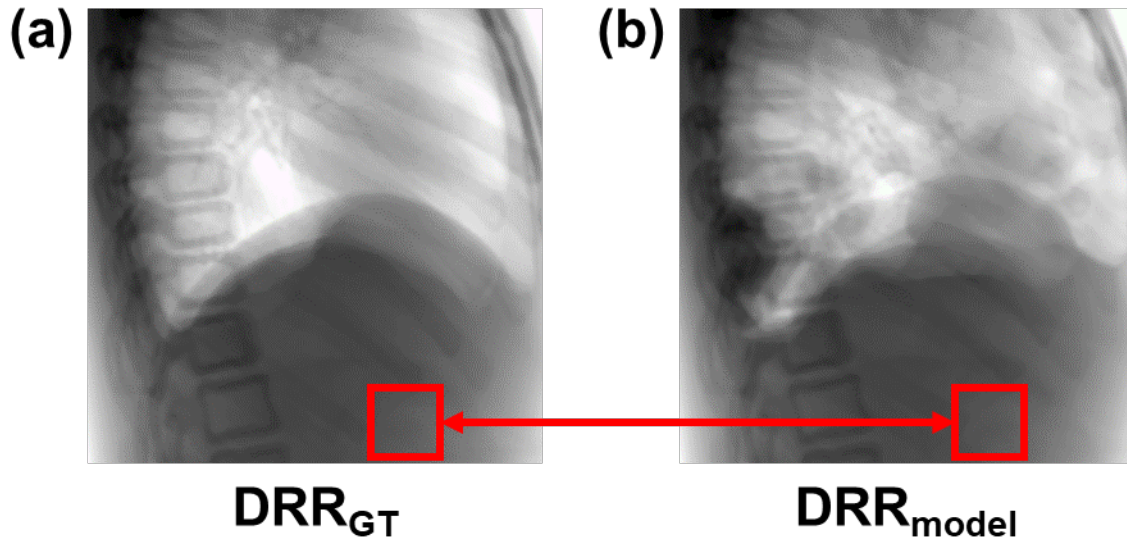


FIGURE 6 DRR images generated from (a) ground truth volume and (b) model volume of the deep breathing T00 phase at an imaging angle 270° . The registration area for model correction is shown in red squares.

To address the errors from the motion model, real-time model correction using the 2D real-time images was carried out. Model correction was performed on a small area at the pancreas region, as shown in Figure 7. As a proof-of-concept study, the area is selected so that the rib movement would not affect the correction process. The registration between the two area generated a correction vector and converted to 3D. The conversion is straightforward as the imaging angle is at 270° and the isocenters of the volume and DRR are well-defined.

Figure 7a shows the ground truth 3D volume with pancreas being highlighted for visualization. Figure 7b and 7c show the difference images between the GT and the motion model results before and after correction, respectively. Visually, the corrected motion model volume shows a reduced error when compared to the GT.

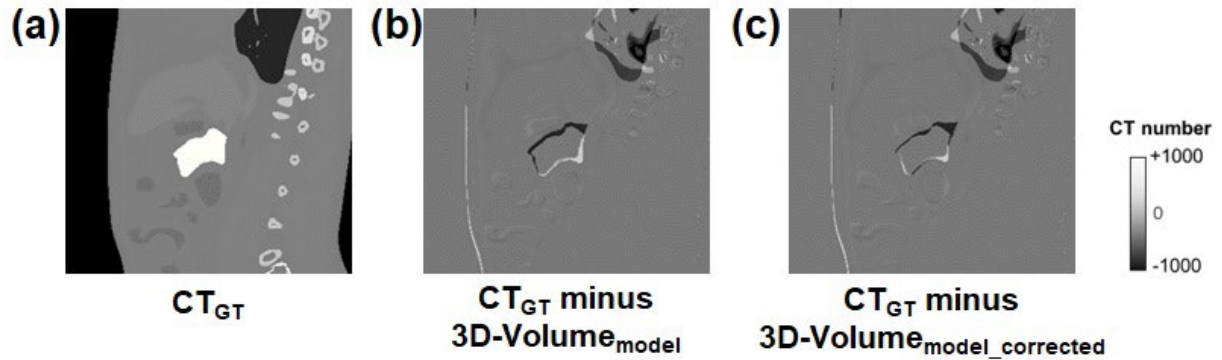


FIGURE 7 (a) CT image (sagittal plane) of the ground truth (GT) created by XCAT and difference images between the GT and motion model results (b) before and (c) after the correction. Pancreas is highlighted for visualization.

Dice coefficients were calculated for pancreas between the GT and motion model volumes with and without correction for the 20 phases. Results show improvement on the Dice coefficient value after the correction process for most phases, as shown in Figure 8. For the fully inhaled state T00, the Dice coefficient value improved from 0.74 to 0.78.

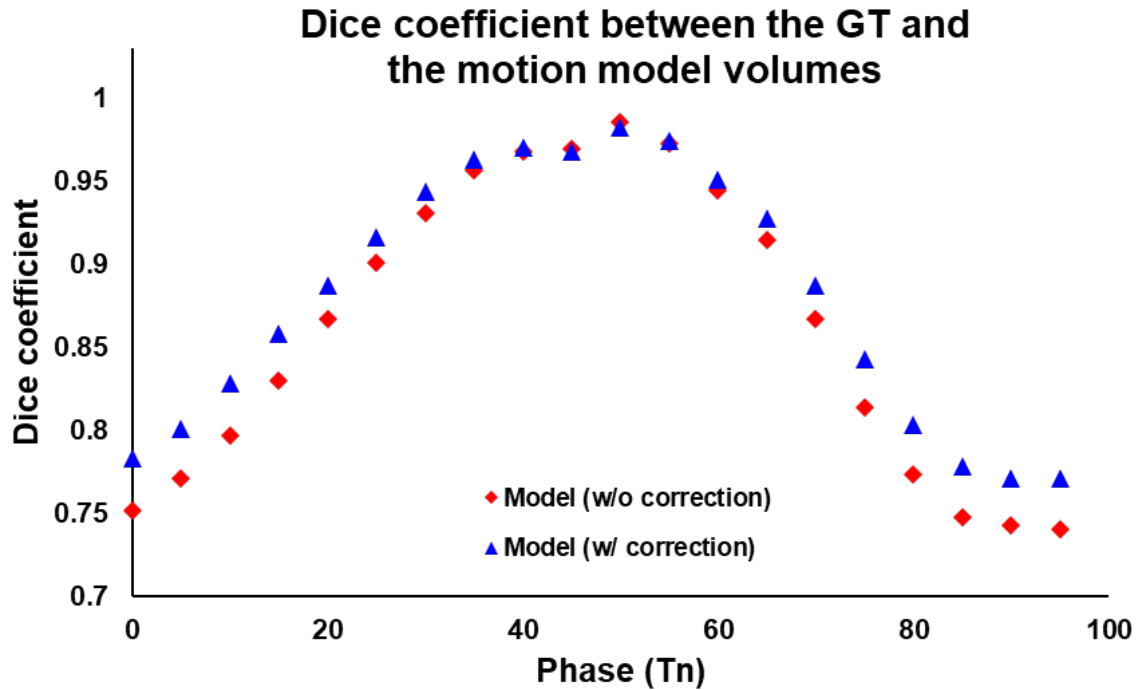


FIGURE 8 Dice coefficients of pancreas between the GT and the motion model volumes without and with correction in digital phantom.

3.2 Physical phantom

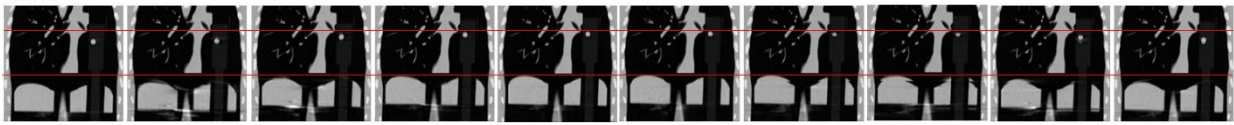
Next, we validate the proposed method experimentally under the condition that is similar to a typical PBT treatment. Figure 9 shows the coronal plane slices for 10 phases of 4DCT images (T00-T90) taken for the phantom which underwent normal breathing motion, for both lung and pancreas environments. DIR was performed between the images of T50 and other phases to obtain a set of DVFs for motion model creation.

The diaphragm top point was tracked in real-time X-ray image pair and converted to 3D displacement, which was used as the input of the motion model. The inherent tracking error for the diaphragm and pancreas tumor was found to be 0.2 mm and 0.5 mm, respectively, measured from the two motion patterns shown in the Supplementary

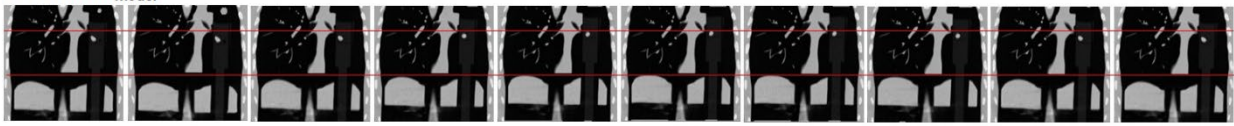
Information. The motion model then generated the 3D volume of each phase. The resulted 3D volumes are also shown in Figure 9. In contrast to the noticeable motion artifacts of diaphragm and rod seen in the experimental 4DCT, the 3D-volume_{model} shows no artifacts as each phase was generated by applying the motion model estimation to the artifact-free reference T50 volume.

(a) Lung tumor simulation

4DCT



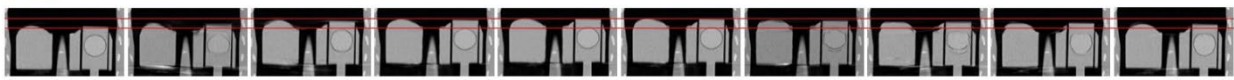
3D-volume_{model}



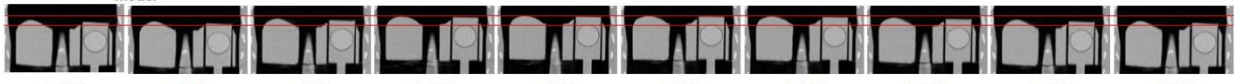
T00 T10 T20 T30 T40 T50 T60 T70 T80 T90

(b) pancreas tumor simulation

4DCT



3D-volume_{model}



T00 T10 T20 T30 T40 T50 T60 T70 T80 T90

FIGURE 9 4DCT images (coronal plane) taken at normal breathing motion and 3D volume generated by motion model using the diaphragm position of the same 10 phases as the input for (a) lung tumor and (b) pancreas tumor simulation. The red lines indicate the positions of tumor and diaphragm at T50 reference state.

The real-time X-ray images at imaging angles of $135^\circ/225^\circ$ of the reference T50 and deep breathing T00 phase, as well as the simulated 2D-Image_{model} from motion model,

are shown in Figure 10. Rigid registrations were carried out in the areas shown in the red rectangles. For the lung case, the area includes the rod, tumor, rib, and mediastinum. For the pancreas case, the area includes the diaphragm, rod, tumor, and the air gap between tumor and rod, while excluding the air gap between the rod and diaphragm. The rigid registrations generate sets of 2D correction vectors. The 2D correction vectors are used to calculate a 3D correction vector using trigonometry. The tumor position is then obtained from the corrected 3D volume.

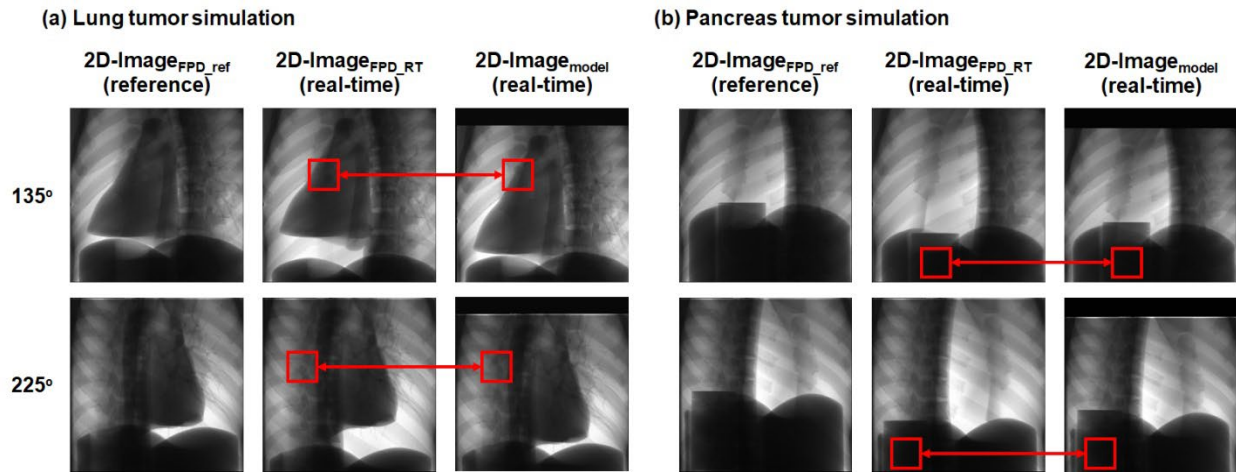


FIGURE 10 Real-time X-ray images at imaging angles of 135°/225° at reference T50 and deep breathing T00 phase of the deep breathing motion, and the corresponding simulated 2D-Image_{model} from motion model, for (a) lung tumor and (b) pancreas tumor simulation. The registration area for model correction is shown in red squares.

Figure 11a compares the ground truth diaphragm and the lung tumor motions, with the measured diaphragm motion and estimated tumor motions from the motion model with and without the real-time correction. The different motions between the diaphragm and lung tumor during the deep breathing condition (motion trace 1) introduce a large error to

the uncorrected motion model estimation. The lung tumor localization errors are shown in Figure 11b. The errors are reduced significantly after applying the correction using the real-time X-ray image. For the fully inhaled states, the tumor localization errors reduce from around 6 mm to around 2 mm after correction, while the mean absolute error (MAE) decreases from 3.3 mm to 0.7 mm after applying the correction. The standard deviations of the error are 2.3 mm before the correction and 0.8 mm after applying the correction. Similar results are found for the shallow breathing condition with a 1.7 mm baseline shift (motion trace 2), as shown in Figure 11c and 11d. The mean absolute error (MAE) decreases from 1.4 mm to 0.8 mm after applying the correction. The standard deviations of the error are 1.0 mm before the correction and 0.5 mm after applying the correction.

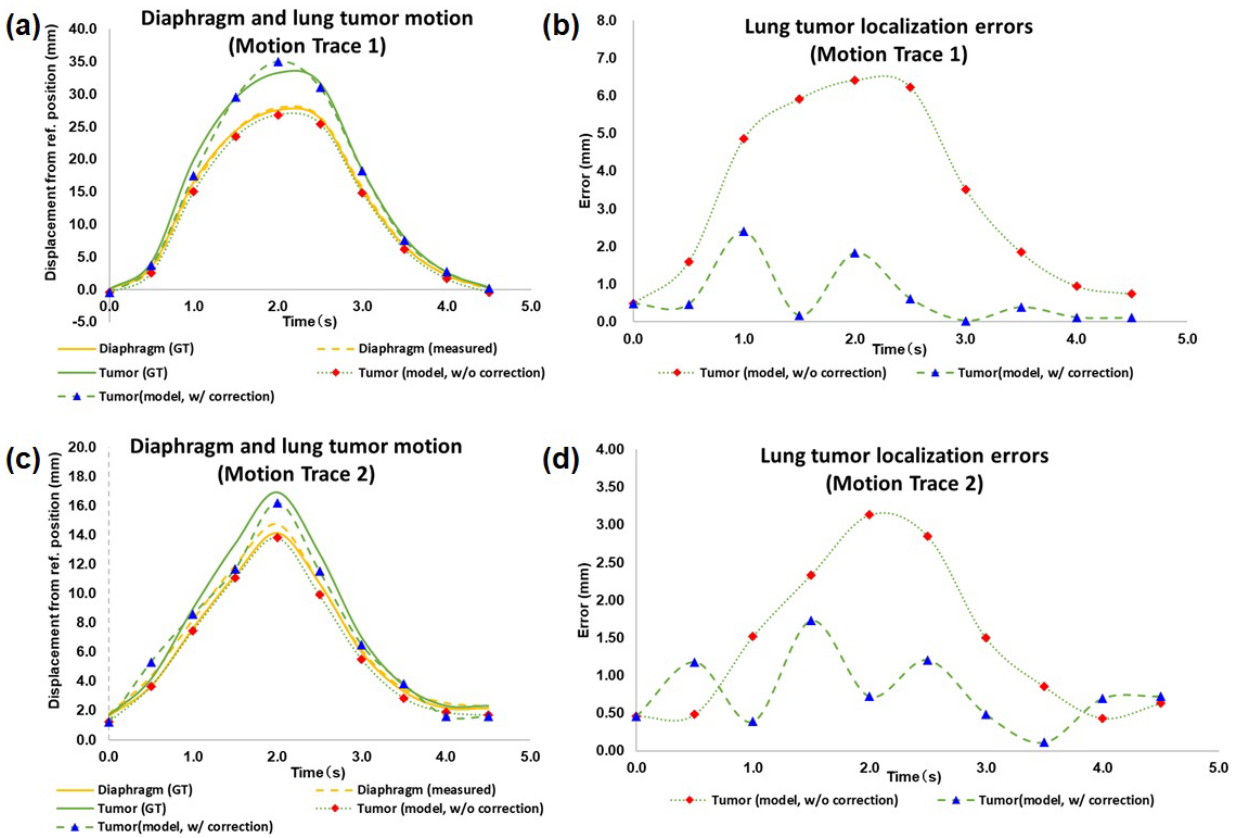


FIGURE 11 Diaphragm and lung tumor motion curves of ground truth (GT) and model without and with correction, Tumor localization errors of the motion model without and with correction, for (a) motion trace 1 and (b) motion trace 2.

Figure 12 shows the similar comparison between the ground truth diaphragm and the pancreas tumor motions, with the measured diaphragm motion and estimated tumor motions from the motion model with and without the real-time correction. The pancreas tumor localization errors are reduced significantly after applying the correction using the real-time X-ray image. For the fully inhaled states during deep breathing (motion trace 1), the tumor localization errors reduce from around 5 mm to less than 2 mm after correction, while the mean absolute error (MAE) decreases from 2.5 mm to 0.5 mm after applying the correction. The standard deviations of the error are 2.1 mm before the correction and 0.5 mm after applying the correction. Similar results are found for the shallow breathing condition with a 1.7 mm baseline shift (motion trace 2), as shown in Figure 11c and 11d. The mean absolute error (MAE) decreases from 1.3 mm to 0.6 mm after applying the correction. The standard deviations of the error are 0.9 mm before the correction and 0.3 mm after applying the correction.

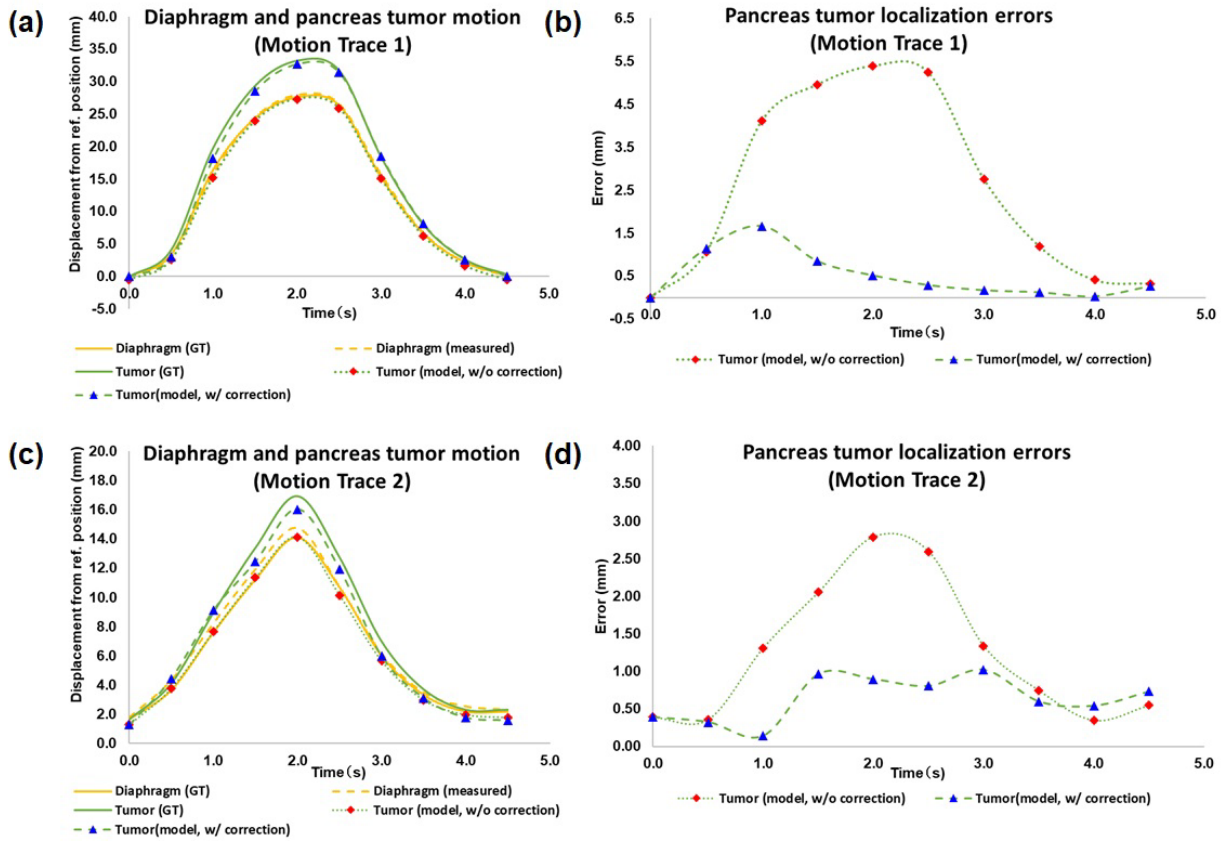


FIGURE 12 Diaphragm and pancreas tumor motion curves of ground truth (GT) and model without and with correction, Tumor localization errors of the motion model without and with correction, for (a) motion trace 1 and (b) motion trace 2.

4 DISCUSSION

A proof-of-concept study for our motion model-based markerless tracking method has been carried out.

In digital phantom, using the diaphragm top point as the input, the motion model generated the DRRs that simulate the instant breathing condition. The registration between the estimated DRR and the ground truth DRR in the pancreas region successfully improved the accuracy of the original estimation from motion model.

Generally, the improvement is significant at larger amplitude cases, where the motion model has larger errors. The improvement is not noticeable or even negative for the exhaled phases where the motion model already has a small error. This is probably due to the registration errors in the correction process due to small displacement.

In physical phantom, the diaphragm top point motion was tracked in X-ray image and was used as the motion model input. The motion model generated 3D and 2D motion estimations, which were used to simulate 2D images with the same modality as the X-ray images. The simulated 2D images were registered to the real-time X-ray images and corrected the original estimated 3D volume. The results show significant reductions of tumor localization error, which is originated from the change in surrogate-tumor relationship between the time when motion model was created and the time during treatment, from around 6 mm to around 2 mm. The success of performing real-time correction to the 3D volume using a pair 2D images allows the tracking accuracy of the motion model to be ensured even when the surrogate-tumor relationship was significantly different than the time when the motion model was created. In terms of computational time, the real-time correction process was performed in around 500 ms using a domestic GPU (NVIDIA GeForce RTX 3050), which allows the implementation at 2 Hz. The majority time spent was on the rigid registration between the 2D images, which was around 400 ms. The creation of real-time DRR_{model} from the real-time 3D-DVF took around 50 ms. While the current proof-of-concept study has a latency of around 500 ms, it is promising for it to be applied in real-time with further optimization of the algorithm in the future.

As a proof-of-concept study, the validation processes were relatively simple and straightforward. For the digital phantom, the CT number of each organ is assumed to be constant across that whole organ. In addition, there is no noise in the DRR images created from digital phantom. For physical phantom, the diaphragm and tumor rod only moved in one direction without any deformation. Moreover, although the tumor rods and the 3D printed materials are suitable for creating a low-contrast environment, it is still far different than the actual environment in patient body. A more accurate simulation and validation using patient data are desired. Finally, all the registrations including the real-time correction was only performed using rigid registration. Improvement of the method to perform deformable real-time correction is needed, such as extending the use of 2D transformation vector into 2D-DVF. However, the conversion from 2D to 3D is an ill-posed problem as there might not be sufficient information to convert 2D deformable vectors to 3D for model correction. As a result, a method to iteratively optimize the 3D real-time volume to simulate a pair of 2D images that match the 2D real-time images needs to be developed.

In addition, our method has the following limitations. The accuracy of our motion model depends on the accuracy of DIR performed between the phases in 4DCT. While the errors can be partly corrected by X-ray image in the real-time correction process, the initial estimation accuracy is still crucial to the correction speed and accuracy. Also, in order to achieve the matching of estimated image and real-time image in the same modality, it is essential to develop a method to establish an accurate correspondence between the reference CT and the X-ray image taken at the reference moment, so that real-time information and the 3D structural information can be connected well. The fore-mentioned

method to iteratively optimize the 3D reference volume that best reproduces a pair of 2D reference images can be one possible option.

Despite the simpleness of our validations and the limitations explained above, our method has successfully combined the detailedness of 4DCT and real-time information from 2D X-ray images through a motion model to track the tumor and tissues motions. In addition, such incorporation was achieved through the matching between the estimated images and real-time images in the same modality, which allows the potential applications in low X-ray contrast regions such as liver and pancreas.

The framework can be generalized in Figure 13. Structural information is modelled into a motion model, and the motion model is corrected, or optimized, by the real-time information to generate 3D volume that can reproduce the real-time information. Such optimization process can be performed iteratively until the similarity index between the estimated images and real-time images reaches a certain threshold. The 3D real-time volume allows the estimation of tumor positions as well as the OAR positions. It also allows the calculation of water equivalent thickness (WET) along the beam path for PBT treatment.

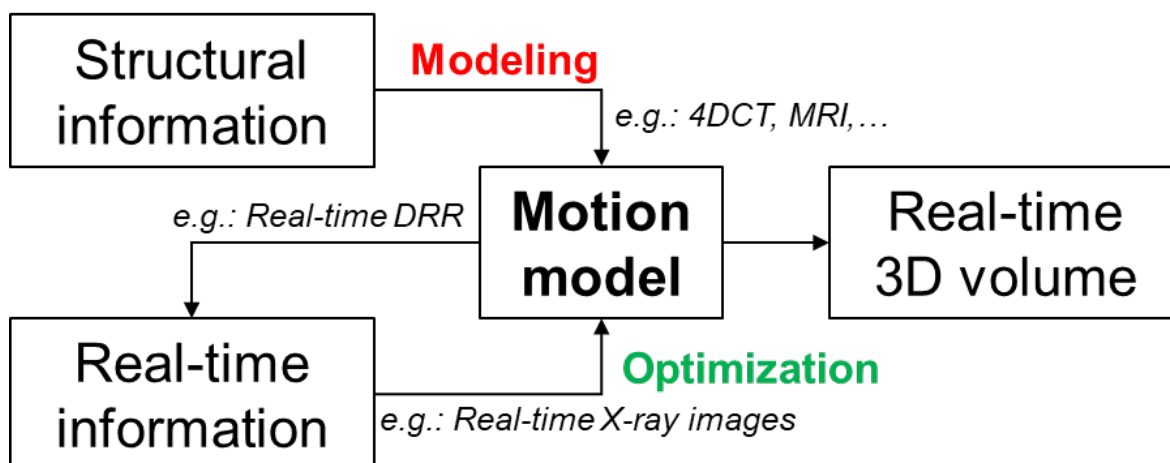


FIGURE 13 The generalized framework of the workflow in this study. Structural information is modelled to create a motion model, while the real-time information is used to optimize the motion model to generate an accurate 3D real-time volume.

Under the framework, our method can be further extended. For example, the structural information can be modelled individually to create an organ model. Such model can be utilized so that real-time monitoring of OAR motion can be achieved, which helps ensure the organs-at-risk (OARs) not to be exposed to excessive dose. A gas model can also be incorporated so that the sudden and quick movement of gas inside the digestive tracts can be tracked. This can be especially useful in PBT as these gas bubbles can significantly alter the beam range of particle beam when they enter the beam path. Moreover, other prior-knowledge and extra freedoms can be used to modify the 3D motion estimation to produce a better matching with the real-time information to account for the deformations of tumor and tissues. For example, some pre-built deformation models such as a biomechanical model can be used as the initial modification.³³ A randomized DVF can be added as a fine-tuning parameter so that the optimization will not be limited by

prior information. This provides the freedom to modify the motion model estimation to describe any possible body conditions appeared during the treatment.

Finally, this framework might also serve as a real-time optimization tool for any 3D real-time estimation result including the AI-based methods. While methods like deep learning are powerful to localize tumor position, it has been difficult to ensure their real-time tracking results are reliable. By incorporating with this framework, AI-based estimations can be further optimized to produce the best simulation that matches the pair of 2D real-time images, thus ensuring its tracking accuracy.

5 CONCLUSION

The key to perform high precision markerless motion tracking of tumor lies on the utilization of detailed structural information from the planning stage and real-time information during treatment. We present a novel approach to incorporate 4DCT and 2D real-time images through a motion model. The proof-of-concept validation results in both digital and physical phantoms show that the 2D real-time images can successfully correct the initial motion model estimation, thus improving the tracking accuracy. Since the matching between the motion model estimation and real-time information is performed between the images in the same modality, tracking is achievable even when the contrasts in real-time X-ray images are low. Our result suggests a new motion model approach to perform markerless tracking.

ACKNOWLEDGMENTS

The authors would like to thank the help from Hokkaido University Hospital for providing supports on the experiment.

CONFLICT OF INTEREST STATEMENT

The authors have no relevant conflicts of interest to disclose.

REFERENCES

1. Mori S, Knopf AC, Umegaki K. Motion management in particle therapy. *Med Phys*. 2018;45:e994-e1010.
2. Bertholet J, Knopf A, Eiben B, et al. Real-time intrafraction motion monitoring in external beam radiotherapy. *Phys Med Biol*. 2019;64:15TR01.
3. Shirato H, Shimizu S, Kunieda T, et al. Physical aspects of a real-time tumor-tracking system for gated radiotherapy. *Int J Radiat Oncol Biol Phys*. 2000;48:1187-1195.
4. Korreman SS. Motion in radiotherapy: photon therapy. *Phys Med Biol*. 2012;57:R161-R191.
5. Mueller M, Poulsen P, Hansen R, et al. The markerless lung target tracking AAPM Grand Challenge (MATCH) results. *Med Phys*. 2022;49:1161-1180.
6. Zhang X, Homma N, Ichiji K, et al. A kernel-based method for markerless tumor tracking in kV fluoroscopic images. *Phys Med Biol*. 2014;59:4897-4911.
7. Ferguson D, Harris T, Shi M, et al. Automated MV markerless tumor tracking for VMAT. *Phys Med Biol*. 2020;65:125011.
8. van Sörnsen de Koste JR, Dahele M, et al. Markerless tracking of small lung tumors for stereotactic radiotherapy. *Med Phys*. 2015;42:1640-1652.
9. Shieh CC, Caillet V, Dunbar M, et al. A Bayesian approach for three-dimensional markerless tumor tracking using kV imaging during lung radiotherapy. *Phys Med Biol*. 2017;62:3065-3080.
10. Cui Y, Dy JG, Sharp GC, et al. Multiple template-based fluoroscopic tracking of lung tumor mass without implanted fiducial markers. *Phys Med Biol*. 2007;52:6229-6242.
11. Kilby W, Dooley JR, Kuduvali G, et al. The CyberKnife Robotic Radiosurgery System in 2010. *Technol Cancer Res Treat*. 2010;9:433-452.

- 599 12. Ziegler M, Lettmaier S, Fietkau R, Bert C. Performance of Markerless Tracking for Gimbaled Dynamic
600 Tumor Tracking. *Z Med Phys.* 2020;30:96-103.
- 601 13. Mori S, Karube M, Shirai T, et al. Carbon-Ion Pencil Beam Scanning Treatment With Gated Markerless
602 Tumor Tracking: An Analysis of Positional Accuracy. *Int J Radiat Oncol Biol Phys.* 2016;95:258-266.
- 603 14. Schnarr E, Beneke M, Casey D, et al. Feasibility of real-time motion management with helical
604 tomotherapy. *Med Phys.* 2018;45:1329-1337.
- 605 15. Zhao W, Shen L, Han B, et al. Markerless Pancreatic Tumor Target Localization Enabled By Deep
606 Learning. *Int J Radiat Oncol Biol Phys.* 2019;105:432-439.
- 607 16. Teo TP, Ahmed SB, Kawalec P, et al. Feasibility of predicting tumor motion using online data acquired
608 during treatment and a generalized neural network optimized with offline patient tumor trajectories. *Med*
609 *Phys.* 201;45:830-845.
- 610 17. Feng M, Balter JM, Normolle D, et al. Characterization of pancreatic tumor motion using cine MRI:
611 surrogates for tumor position should be used with caution. *Int J Radiat Oncol Biol Phys.* 2009;74:884-891.
- 612 18. Bert C, Durante M. Motion in radiotherapy: particle therapy. *Phys Med Biol.* 2011;56:R113-144.
- 613 19. McClelland JR, Hawkes DJ, Schaeffter T, King AP. Respiratory motion models: a review. *Med Image*
614 *Anal.* 2013;17:19-42.
- 615 20. Fassi A, Seregini M, Riboldi M, et al. Surrogate-driven deformable motion model for organ motion
616 tracking in particle radiation therapy. *Phys Med Biol.* 2015;60:1565-1582.
- 617 21. Fassi A, Schaerer J, Fernandes M, et al. Tumor tracking method based on a deformable 4D CT
618 breathing motion model driven by an external surface surrogate. *Int J Radiat Oncol Biol Phys.* 2014;88:182-
619 188.
- 620 22. Fayad H, Pan T, Pradier O, Visvikis D. Patient specific respiratory motion modeling using a 3D patient's
621 external surface. *Med Phys.* 2012;39:3386-3395.
- 622 23. Li R, Jia X, Lewis JH, et al. Real-time volumetric image reconstruction and 3D tumor localization based
623 on a single x-ray projection image for lung cancer radiotherapy. *Med Phys.* 2010;37:2822-2826.
- 624 24. Li R, Lewis JH, Jia X, et al. 3D tumor localization through real-time volumetric x-ray imaging for lung
625 cancer radiotherapy. *Med Phys.* 2011;38:2783-2794.

- 626 25. Asano S, Oseki K, Takao S, et al. Technical note: Performance evaluation of volumetric imaging based
627 on motion modeling by principal component analysis. *Med Phys*. 2023;50:993-999.
- 628 26. Garau N, Via R, Meschini G, et al. A ROI-based global motion model established on 4DCT and 2D cine-
629 MRI data for MRI-guidance in radiation therapy. *Phys Med Biol*. 2019;64:045002.
- 630 27. Dasnoy-Sumell D, Aspeel A, Souris K, Macq B. Locally tuned deformation fields combination for 2D
631 cine-MRI-based driving of 3D motion models. *Phys Med*. 2022;94:8-16.
- 632 28. Hindley N, Keall P, Booth J, Shieh CC. Real-time direct diaphragm tracking using kV imaging on a
633 standard linear accelerator. *Med Phys*. 2019;46:4481-4489.
- 634 29. Bahig H, Campeau MP, Vu T, et al. Predictive parameters of CyberKnife fiducial-less (XSight Lung)
635 applicability for treatment of early non-small cell lung cancer: a single-center experience. *Int J Radiat Oncol*
636 *Biol Phys*. 2013;87:583-589.
- 637 30. Segars WP, Sturgeon G, Mendonca S, Grimes J, Tsui BM. 4D XCAT phantom for multimodality imaging
638 research. *Med Phys*. 2010;37:4902-4915.
- 639 31. Sharp GC, Li R, Wolfgang J, et al. Plastimatch: an open source software suite for radiotherapy image
640 processing. *Proceedings of the XVI'th International Conference on the use of Computers in Radiotherapy*
641 *(ICCR)*, Amsterdam, Netherlands. 2010.
- 642 32. Ding GX, Alaei P, Curran B, , et al. Image guidance doses delivered during radiotherapy: Quantification,
643 management, and reduction: Report of the AAPM Therapy Physics Committee Task Group 180. *Med Phys*.
644 2018 May;45:e84-e99.
- 645 33. Shao HC, Wang J, Bai T, et al. Real-time liver tumor localization via a single x-ray projection using deep
646 graph neural network-assisted biomechanical modeling. *Phys Med Biol*. 2022;67:115009.