Title:

Real-time CT image generation based on voxel-by-voxel modeling of internal deformation by utilizing the displacement of fiducial markers

Short running title:

Real-time CT image generation

Authors:

Risa Hayashi[1], Koichi Miyazaki[2,3], Seishin Takao[2,3], Kohei Yokokawa[2], Sodai Tanaka[2,3], Taeko Matsuura[2,3], Hiroshi Taguchi[4], Norio Katoh[5], Shinichi Shimizu[3,5], Kikuo Umegaki[2,3], Naoki Miyamoto[2,3,*]

[1] Graduate School of Engineering, Hokkaido University, North 13, West 8, Kita-ku, Sapporo, Hokkaido, 060-8628, Japan

[2] Faculty of Engineering, Hokkaido University, North13, West 8, Kita-ku, Sapporo, Hokkaido, 060-8628, Japan

[3] Department of Medical Physics, Hokkaido University Hospital, North 14, West 5, Kita-ku, Sapporo, Hokkaido, 060-8648, Japan

[4] Department of Radiation Oncology, Hokkaido University Hospital, North 14, West 5, Kita-ku, Sapporo, Hokkaido, 060-8648, Japan

[5] Faculty of Medicine, Hokkaido University, North 15, West 7, Kita-ku, Sapporo, Hokkaido, 060-8638, Japan

Corresponding author:

Naoki Miyamoto,

Faculty of Engineering, Hokkaido University, North 13, West 8, Kita-ku, Sapporo, Hokkaido, 060-8628, Japan.

Tel: +81-11-706-6673

E-mail address: miya-nao@eng.hokudai.ac.jp

Abstract

Purpose

To show the feasibility of real-time CT image generation technique utilizing internal fiducial markers that facilitate the evaluation of internal deformation.

Methods

In the proposed method, linear regression model that can derive internal deformation from the displacement of fiducial markers is built for each voxel in the training process before the treatment session. Marker displacement and internal deformation are derived from the four-dimensional computed tomography (4DCT) dataset. In the treatment session, the three-dimensional deformation vector field is derived according to the marker displacement, which is monitored by the real-time imaging system. The whole CT image can be synthesized by deforming the reference CT image with a deformation vector field in real-time. To show the feasibility of the technique, image synthesis accuracy, and tumor localization accuracy were evaluated using dataset generated by extended NURBS-Based Cardiac-Torso (XCAT) phantom and clinical 4DCT datasets from six patients, containing ten CT datasets each. In the validation with XCAT phantom, motion range of the tumor in training data and validation data were about 10 and 15 mm, respectively, so as to simulate motion variation between 4DCT acquisition and treatment session. In the validation with patient 4DCT dataset, eight CT datasets from the 4DCT dataset were used in the training process. Two excluded inhale CT datasets can be regarded as the datasets with large deformations more than training dataset. CT images were generated for each respiratory phase using the corresponding marker displacement. Root mean squared error (RMSE), normalized RMSE (NRMSE) and structural similarity index measure (SSIM) between the original CT images and the synthesized CT images were evaluated as the quantitative indices of the accuracy of image synthesis. The accuracy of tumor localization was also evaluated.

Results

In the validation with XCAT phantom, the mean NRMSE, SSIM and three-dimensional tumor localization error were 7.5±1.1%, 0.95±0.02 and 0.4±0.3 mm, respectively. In the validation with patient 4DCT dataset, the mean RMSE, NRMSE, SSIM and three-dimensional tumor localization error in six patients were 73.7±19.6 HU, 9.2±2.6%, 0.88±0.04 and 0.8±0.6 mm, respectively. These results suggest that the accuracy of the proposed technique is adequate when the respiratory motion is within the range of the training dataset. In the evaluation with a marker displacement larger than that of the training dataset, the mean RMSE, NRMSE, and tumor localization error were about 100 HU, 13%, and less than 2.0 mm, respectively, except for one case having large motion variation. The performance of the proposed method was similar to those of previous studies. Processing time to generate the volumetric image was less than 100 ms.

Conclusion

We have shown the feasibility of the real-time CT image generation technique for volumetric imaging.

Keywords: CT image generation, synthetic CT, volumetric imaging, motion management, fiducial markers, partial least squares regression

Introduction

Respiratory motion is a major source of dosimetric error in radiation therapy. Respiratory-gated irradiation using an external surrogate, such as abdominal motion, has been clinically applied for motion management.[1-5] In addition, real-time tumor-tracking radiation therapy (RTRT) [6], which monitors the internal tumor motion directly, has been applied in lung[7,8], liver[9] and pancreas. In RTRT, the three-dimensional position of internal fiducial markers, which are inserted near the tumor, are monitored in real-time by means of orthogonal x-ray imaging during treatment. Treatment beam is irradiated only when the fiducial marker is within the gating window, which is cubic region of $\pm$2 mm in typical clinical setting. RTRT has been also clinically applied in proton therapy[10,11]. Tracking irradiation techniques that change the irradiation position/field according to the tumor position by using electromagnetic transponder guided multileaf collimator (MLC)[12], a compact linear accelerator mounted on a robotic arm[13] and the gimbaled x-ray head[14] have also been realized in photon therapy. However, for accurate irradiation, especially in particle therapy and stereotactic body radiation therapy (SBRT), both the tumor position and volumetric information of the surrounding organs and tissue are needed for real-time motion management.

Real-time volumetric imaging, which can reconstruct anatomical structures, could be applied for a novel motion management technique. Volumetric imaging, based on the estimation of the coefficients of the principal components of internal deformation from a planar x-ray image, has been studied.[15,16] Other approaches based on deep learning have also been reported.[17-19] Previous studies demonstrated that clinically acceptable accuracy of image synthesis and tumor localization can be obtained. However, imaging accuracy may depend on imaging angle and quality.[19]

In this study, we propose a novel real-time CT image generation technique for volumetric imaging by utilizing the positional displacement of internal fiducial markers evaluated in real-time during treatment. In RTRT using an orthogonal x-ray imaging system, the three-dimensional position of the fiducial markers can be obtained with a high spatial accuracy of less than 1 mm and a high temporal resolution up to 30 times per second. Accuracy does not depend as much on image quality of planar x-rays, as long as the marker can be recognized correctly.[20] In addition, doses due to x-ray imaging can be suppressed by collimating the imaging area using only the regions of the fiducial

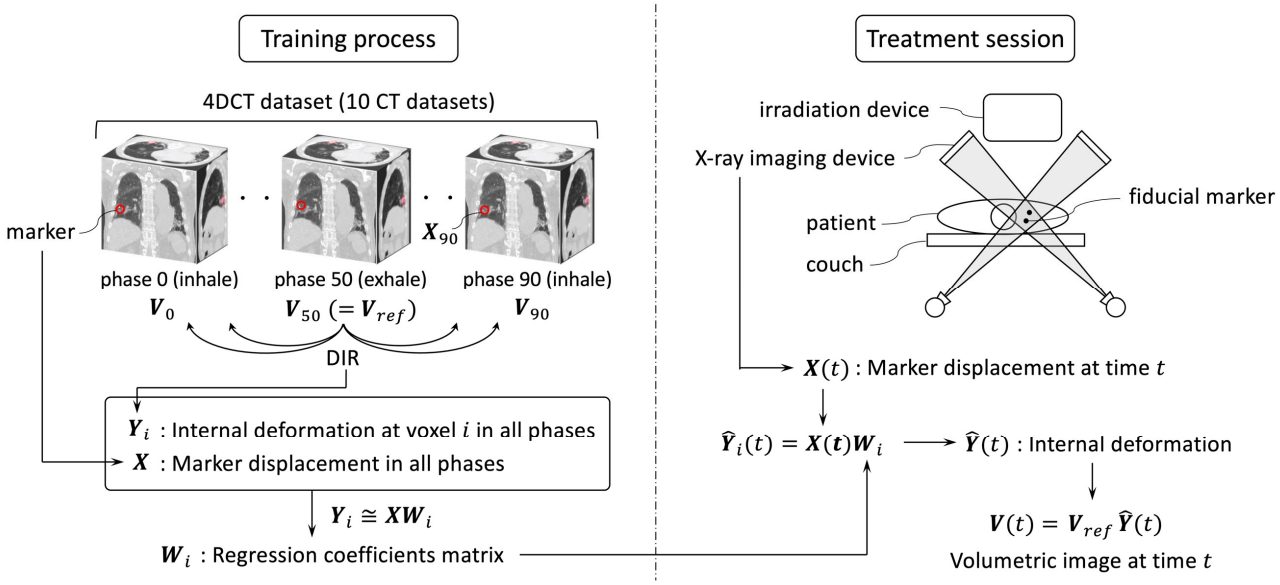
markers. The positional displacement of the internal fiducial markers, which are inserted near the tumor, should be an appropriate surrogate to evaluate internal deformation. Fewer studies focus on volumetric imaging based on internal fiducial markers. In the proposed method, linear regression coefficients to derive the three-dimensional internal deformation from the displacement of the fiducial markers are optimized in each voxel using four-dimensional computed tomography (4DCT) images as the training dataset. CT image can be generated in real-time by applying the deformation derived by the displacement of the fiducial markers to the reference CT image. The purpose of this study was to demonstrate the clinical feasibility of the proposed real-time CT image generation technique. The accuracy of image synthesis and tumor localization were evaluated using dataset generated by extended NURBS-Based Cardiac-Torso (XCAT[21]) phantom and clinical 4DCT datasets from six patients.

Materials and Methods

Algorithm of real-time CT image generation

Figure 1 shows the general scheme of the proposed real-time CT image generation technique, including the training process and the assumed use in the treatment session. In the training process (before the treatment), a linear regression coefficients matrix to evaluate the internal deformation from the displacement of the fiducial markers is generated for each voxel using a 4DCT dataset. The internal deformation $\boldsymbol{Y}_i$ is derived from the deformable image registration (DIR) relative to the exhale phase CT, referred to as phase 50 in this study, as the reference volume $\boldsymbol{V}_{ref}$. $\boldsymbol{Y}_i$ is a matrix with three columns and the number of rows for the matrix equals the number of phases of 4DCT, including the three-dimensional deformation in all respiratory phases at voxel i. The displacement of the fiducial markers $\boldsymbol{X}$ is evaluated for each CT dataset relative to the reference. $\boldsymbol{X}$ is a matrix whose columns and rows are $3n$ and the number of phases, respectively, including the three-dimensional displacement of the n internal fiducial markers in all respiratory phases. Assuming a linear relationship between the internal deformation and the displacement of the fiducial markers, their relationship is described as $\boldsymbol{Y}_i \cong \boldsymbol{X}\boldsymbol{W}_i$, where $\boldsymbol{W}_i$ is a matrix of linear regression coefficients whose columns and rows are 3 and $3n$, respectively. The linear relationship in the lung region, including the tumor, was considered to be reasonable since the fiducial markers were inserted around the tumor. $\boldsymbol{W}_i$ can be derived mathematically by regarding $\boldsymbol{X}$ and $\boldsymbol{Y}_i$ as the explanatory and response variables,

111 respectively. When there is a high correlation between the explanatory variables, output values may fluctuate
112 significantly due to multicollinearity in the ordinary multiple linear regression. In this study, partial least squares
113 regression (PLSR)[22,23], which can address multicollinearity in regression analysis, was utilized to derive $\boldsymbol{W}_i$.



114
115
116 **Fig. 1**. General scheme of the proposed real-time CT image generation technique.
117

118 In the first step of the treatment session, a patient is aligned to the planning position, according to rigid anatomical
119 landmarks such as the vertebra, by an image guidance technique, such as cone beam CT (CBCT). In this study, we
120 assumed an orthogonal x-ray imaging system to obtain the three-dimensional position of the internal fiducial
121 markers. Treatment system with floor/ceiling mounted orthogonal imaging device has been clinically applied in
122 photon therapy[6], and gantry mounted type has been realized in proton therapy[10]. Please refer the literatures for the
123 detail of treatment system. Geometrical relationship between the iso-center and the reference position of the fiducial
124 markers is determined in treatment planning process in advance. The three-dimensional positions of the fiducial
125 markers are obtained in real-time, using orthogonal x-ray imaging during treatment. The three-dimensional
126 positional displacement of the n fiducial markers $\boldsymbol{X}(t)$ at time t is obtained as the difference between the current
127 and reference position of the fiducial markers. The three-dimensional deformation vector $\hat{\boldsymbol{Y}}_i(t)$ at voxel i is derived

as $\widehat{\boldsymbol{Y}}_i(t) = \boldsymbol{X}(\boldsymbol{t})\boldsymbol{W}_i$, according to the marker displacement. The volumetric image $\boldsymbol{V}(t)$ can be synthesized by deforming the reference volume $\boldsymbol{V}_{ref}$ with the deformation vector field $\widehat{\boldsymbol{Y}}(t)$, which is constructed from all voxels.

Validation with XCAT phantom

For the evaluations, a ground truth volumetric image was necessary. In this study, the evaluations were conducted by using the dataset of volumetric images generated by extended NURBS-Based Cardiac-Torso (XCAT) phantom[21] as with the previously reported studies[24,25]. XCAT phantom can generate three-dimensional imaging data according to the arbitrary anterior-posterior (AP) motion data and superior-inferior (SI) motion data for simulating the motion of chest wall and diaphragm, respectively. In this study, in order to mimic actual patient motion, internal marker motion data and external surrogate motion data that were acquired simultaneously by using RTRT system were used. The external surrogate motion and SI direction motion of the fiducial marker were resampled into 5 Hz and were used as input motion data for simulating chest wall and diaphragm motion, respectively, in XCAT phantom. One motion dataset that can simulate the change of motion amplitude between training session and validation session was used. Motion range of the tumor in training session and validation session was about 10 and 15 mm, respectively. Tumor of which diameter was 20 mm and one fiducial marker were inserted in right lower lung region. Image synthesis accuracy and tumor localization accuracy were evaluated as described later.

Validation with patient 4DCT dataset

The evaluations were also conducted with the patient 4DCT dataset. The schematic representation of the testing protocol with patient 4DCT dataset for this study is shown in Figure 2. The 4DCT dataset included ten respiratory phases. Eight CT images from phase 10 to phase 80 were used to obtain the matrix of linear regression coefficients in the training process. The CT images were generated for each respiratory phase using the corresponding marker displacement that was derived from 4DCT dataset and compared with the original CT datasets. The two excluded inhale CT datasets, defining phases 0 and 90, corresponded to validation of the datasets that had large deformations not included in the training process. Evaluation with the training CT datasets containing phases 10 to 80 was regarded as the validation of model goodness of fit.

MATLAB (MathWorks Inc, USA) was used for data processing, including CT image import/export, PLSR, and DIR.

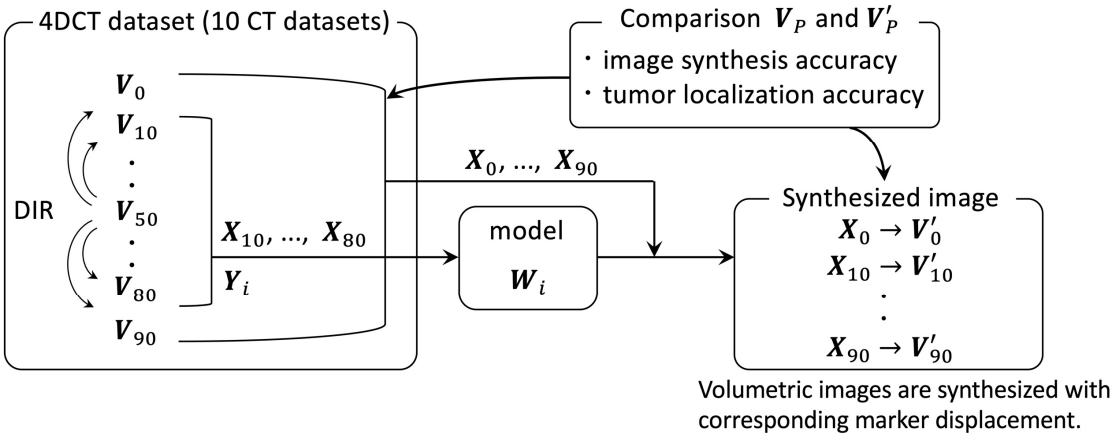


Fig. 2 Schematic representation of the testing protocol with patient 4DCT dataset.

Patient data for evaluation

The 4DCT datasets from six patients, who underwent lung RTRT, were used for the analysis. Slice thickness, image size, and pixel spacing for all CT images were 2.5 mm, 512×512 pixels, and 0.98 mm/pixel, respectively. We set two criteria for data selection. The first criterion was that fiducial markers, with a three-dimensional motion range in ten respiratory phases of more than 5 mm, were included. The second criterion was that the difference in motion range between the ten respiratory phases and the eight respiratory phases without the two inhale phases was more than 1 mm. The motion ranges for each fiducial marker in the six patients are summarized in Table 1. Three or four fiducial markers have been included for each patient. Harada et al. reported[26] that the mean motion amplitude during radiation therapy was larger than that evaluated in a 4DCT dataset by about 2 mm. Four 4DCT datasets in this study had motion differences of more than 2 mm; hence, these evaluations adequately mimicked the clinical situation.

Table 1[1]**.** Motion range for each internal fiducial marker in six patients.

	Motion range in LR/SI/AP/3D [mm]

[1] RML, right middle lobe; LUL, left upper lobe; LLL, left lower lobe; RLL, right lower lobe.

Pat. #	Tumor location	Marker #	from phase 0 to phase 90	from phase 10 to phase 80	difference
#1	RML	1	3.1/7.5/5.7/9.9	2.1/5.5/5.7/8.2	1.0/2.0/0.0/1.7
		2	2.4/10.6/3.3/11.4	2.4/10.6/3.3/11.4	0.0/0.0/0.0/0.0
		3	2.6/9.3/4.6/10.7	1.5/6.5/4.6/ 8.1	1.1/2.8/0.0/2.6
#2	LUL	1	1.8/5.2/2.9/6.2	1.5/4.7/2.7/5.6	0.3/0.5/0.2/0.6
		2	2.6/6.6/3.1/7.7	1.7/5.7/2.5/6.4	0.9/0.9/0.6/1.3
		3	1.2/6.0/3.6/7.1	1.2/4.8/2.9/5.7	0.0/1.2/0.7/1.4
		4	2.4/5.5/5.5/8.1	1.1/5.5/3.8/6.8	1.3/0.0/1.7/1.3
#3	LLL	1	3.2/20.0/6.8/21.4	2.6/10.0/6.8/12.3	0.6/10.1/0.0/9.1
		2	2.2/18.5/4.7/19.3	2.2/10.7/4.7/11.9	0.0/7.8/0.0/7.3
		3	5.0/23.0/9.5/25.4	5.0/12.5/7.0/15.2	0.0/10.5/2.5/10.2
		4	7.2/25.6/5.7/27.2	7.0/22.1/5.1/23.7	0.2/3.5/0.6/3.5
#4	LUL	1	1.6/2.6/2.2/3.7	1.3/2.5/2.0/3.4	0.3/0.1/0.2/0.3
		2	1.5/5.8/2.4/6.4	1.2/4.6/2.4/5.3	0.3/1.2/0.0/1.1
		3	1.9/7.6/2.3/8.2	1.9/6.0/2.3/6.7	0.0/1.6/0.0/1.5
#5	RLL	1	1.8/12.6/3.4/13.1	1.8/7.9/2.0/8.4	0.0/4.7/1.4/4.7
		2	1.8/15.9/2.5/16.2	1.3/11.2/1.4/11.3	0.5/4.7/1.1/4.9
		3	1.4/18.0/3.1/18.4	0.7/12.0/3.0/12.3	0.7/6.0/0.1/6.1
		4	0.5/18.2/1.2/18.2	0.5/17.2/1.2/17.3	0.0/1.0/0.0/0.9
#6	RML	1	4.2/3.2/4.0/6.6	3.6/3.2/3.5/6.0	0.6/0.0/0.5/0.6
		2	1.9/6.0/2.9/6.9	1.9/6.0/2.3/6.7	0.0/0.0/0.6/0.2
		3	2.9/3.2/5.2/6.7	2.2/3.2/2.5/4.6	0.7/0.0/2.7/2.1
		4	4.8/2.5/1.4/5.6	4.2/2.5/1.4/5.0	0.6/0.0/0.0/0.6

Three-dimensional positions of the fiducial markers in the CT images were determined using projection images created in the anterior-posterior (AP) and left-right (LR) directions. First, maximum intensity projection (MIP) images were created. Then, the images were converted to binary images with a threshold CT value of 2000 HU to segment the marker. The marker position was designated as the center of gravity of the rectangular-shaped region of interest enclosing the segmented fiducial marker.

Accuracy of image synthesis

In this study, RMSE and NRMSE between the ground truth CT images and the synthesized CT images were evaluated as the quantitative indices of the accuracy for the image synthesis. RMSE and NRMSE were evaluated as:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(I_i - I_i^*\right)^2},$$

$$\mathrm{NRMSE} = \sqrt{\frac{\sum_{i=1}^{n}\left(I_i - I_i^*\right)^2}{\sum_{i=1}^{n} I_i^{*2}}},$$

where I_i^* is the actual CT value at voxel i in the ground truth image, I_i is the CT value at voxel i in the synthesized image, and n is voxel number. In addition, structural similarity index measure (SSIM) which is commonly used in pattern matching as a measure of the similarity of image structures was evaluated as:

$$\mathrm{SSIM} = \frac{\left(2\mu_x\mu_y + C_1\right)\left(2\sigma_{xy} + C_2\right)}{\left(\mu_x^2 + \mu_y^2 + C_1\right)\left(\sigma_x^2 + \sigma_y^2 + C_2\right)},$$

where μ_x and μ_y are the average CT values in ground truth image and synthesized image respectively, σ_x and σ_y are variance of CT value in ground truth image and synthesized image respectively, and σ_{xy} is a covariance of ground truth image and synthesized image. C_1 and C_2 were defined as $C_1 = (0.01 \times L)^2$ and $C_2 = (0.03 \times L)^2$, respectively. L was dynamic range of the image and was defined as the maximum CT value except for the fiducial marker in the ground truth image. RMSE, NRMSE and SSIM were evaluated in the three-dimensional region constructed from the body half including the tumor and the slices encompassing tumor motion in the superior-inferior (SI) direction. The performance of the proposed technique was limited by the accuracy of the DIR. For the benchmark of the image synthesis accuracy, the reference image, referred to as phase 50 in this study, was deformed to the ground truth images. RMSE, NRMSE and SSIM were evaluated by comparing the ground truth image and the registered image. To quantitatively confirm the goodness of fit in the linear regression, percent variance explained (PVE), which corresponds to the determination coefficient in the linear regression, was examined in the

196 PLSR. In XCAT phantom, NRMSE and SSIM were evaluated since the dataset generated by XCAT phantom
197 consisted of attenuation coefficient, not of CT value.

198 ***Accuracy of tumor localization***

199 Tumor was manually segmented with referring the images binarized with CT value which was evaluated from the
200 tumor in the ground truth image as threshold. After delineation of the tumor region in both ground truth image and
201 the synthetic image, the distance of center-of-mass between two tumor regions was evaluated as tumor localization
202 accuracy. Dice coefficient between two delineated tumor regions was also evaluated in order to confirm the
203 similarity of the tumor shape. Since the uncertainty of tumor delineation in patients #2 and 4 whose tumor in lung
204 peripheral region was in contact with the pleura would be large, evaluation of tumor localization accuracy was
205 limited to four patients, #1, 3, 5, and 6, whose tumor was located in the central lung region.

206 # Results

207 ***Validation with XCAT phantom***

208 Figure 3 shows NRMSE, SSIM, three-dimensional tumor localization error and Dice coefficient along with the
209 tumor motion. All metrics showed better performance when the displacement of tumor position from the reference
210 position was decreased. Relatively large error was observed when the tumor motion exceeded the range of training
211 data. Average of NRMSE, SSIM and three-dimensional tumor localization error were 7.5±1.1% (ranging from 2.2
212 to 10.1%), 0.95±0.02 (ranging from 0.91 to 0.99) and 0.4±0.3 mm (ranging from 0.0 to 1.3 mm), respectively. Trend
213 of NRMSE and SSIM of the synthesized CT images were similar to those of registered images by DIR which can
214 be regarded as benchmark. Examples of ground truth, synthesized, and absolute difference images in the worst and
215 typical NRMSE are shown in Figure 4. Although the difference around the tumor and diaphragm were large in the
216 worst case compared with the typical cases, internal anatomical structures were reasonably reconstructed with
217 NRMSE less than 10%.

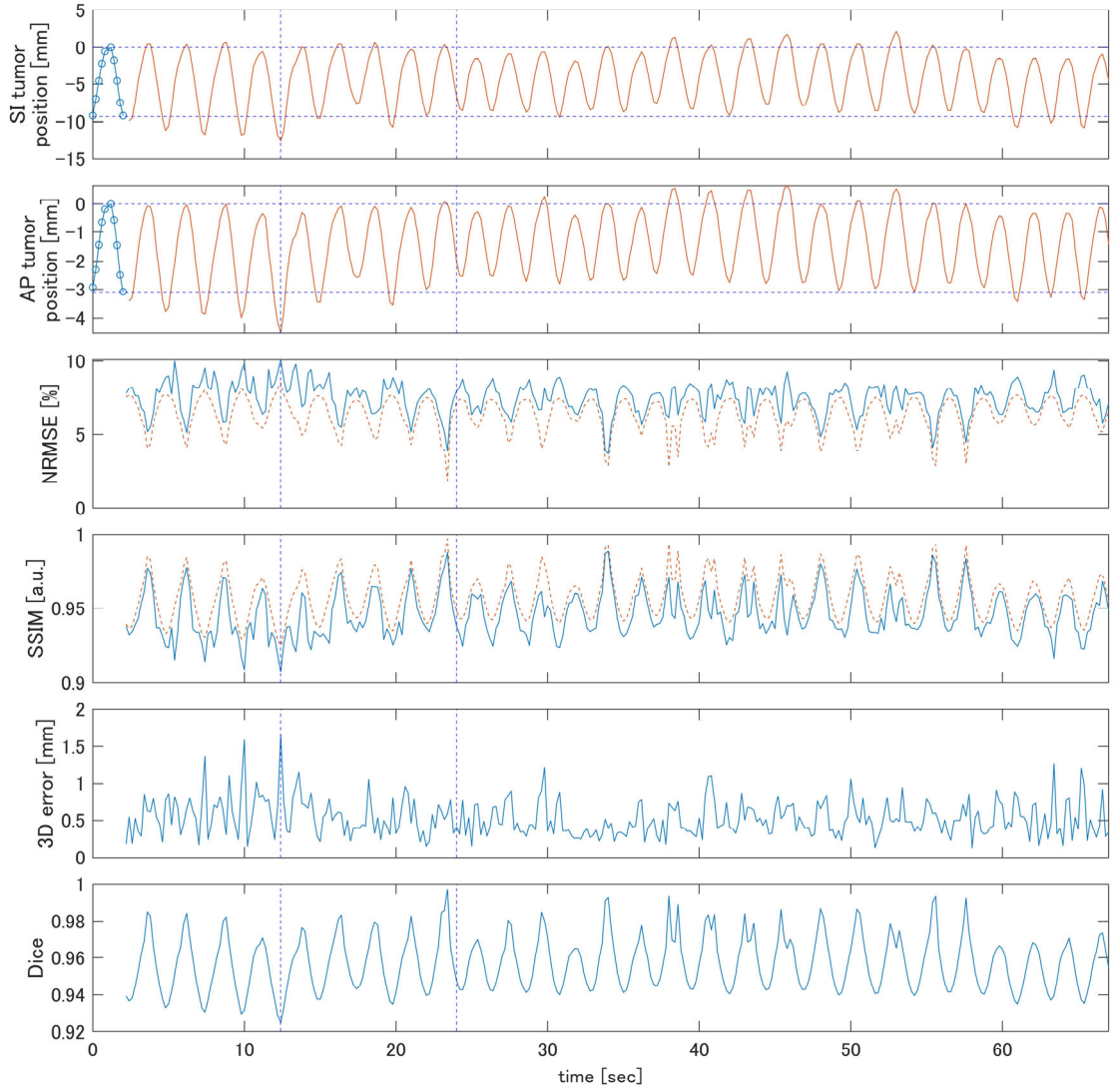


Fig. 3 Top two rows: SI and AP motion of the tumor, respectively. Ten circled points in the first part of the trajectory were corresponding to the training data. The sixth dataset at the exhale phase was used as the reference dataset. The rest data was used for the validation. Horizontal dashed lines show the motion range of the training data. Third row from the top: NRMSE in the validation data. Solid line and dashed line represent the NRMSE evaluated with the synthesized CT images and the registered CT images by DIR, respectively. Fourth row from the top: SSIM in the

validation data. Solid line and dashed line represent the SSIM evaluated with the synthesized CT images and the registered CT images by DIR, respectively. Fifth row from the top: three-dimensional tumor localization error. Bottom row: Dice coefficient between the tumor regions in synthesized images and ground truth images. Vertical dashed lines around 12 sec and 24 sec are corresponding to the data positions having worst and typical NRMSE. The example images in each dataset are shown in Figure 4.

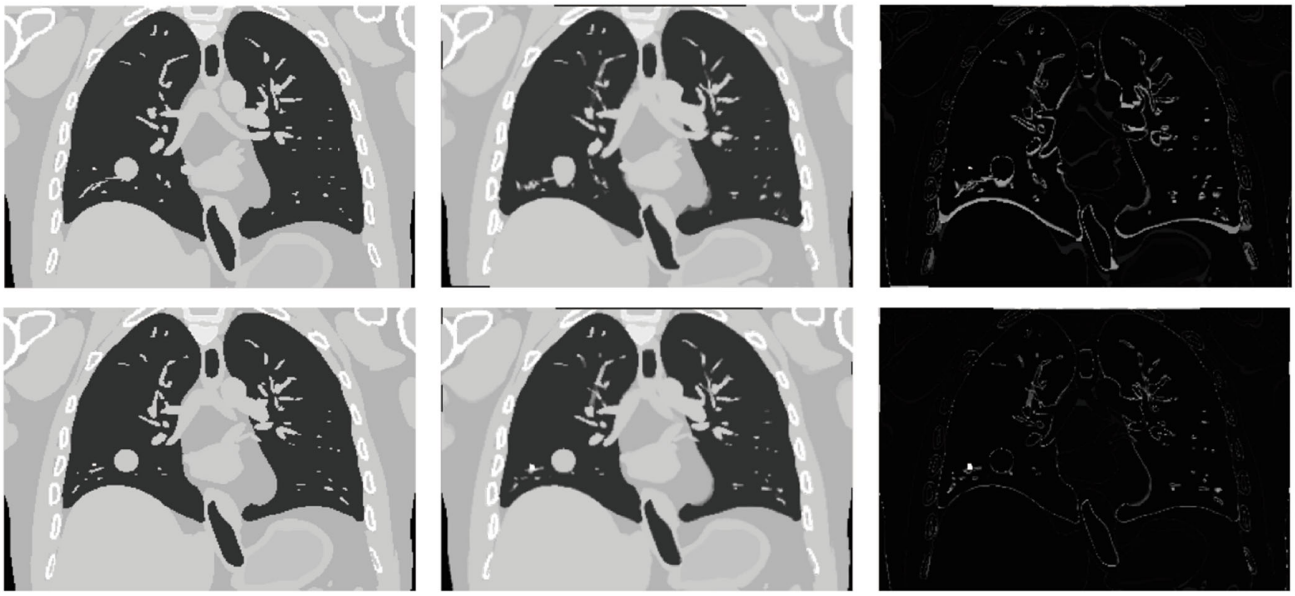


Fig. 4 Top row: from left column, sagittal view of the ground truth image, synthesized image and the absolute difference image between the ground truth image and synthesize image in case of the worst RMSE of 10.1%. Bottom row: images in case of the RMSE of 7.5%. The tumor is a round object in right lower lung region. Fiducial marker is shown at the right side of the tumor.

Accuracy of image synthesis in patient dataset

Examples of MIP images and binary images for the marker segmentation of patient #1 are shown in Figure 5. The rectangle enclosing the segmented marker area was larger than the actual marker size, especially in the SI direction, due to motion artifact. The center of the rectangle was defined as the marker position. In this study, the mean size

of the rectangle in the SI direction was about 5.0 mm. Considering the marker diameter of 1.5 mm, the range of respiratory motion, and CT image acquisition time of several hundred milliseconds, it was thought that most of the marker positions were correctly determined.

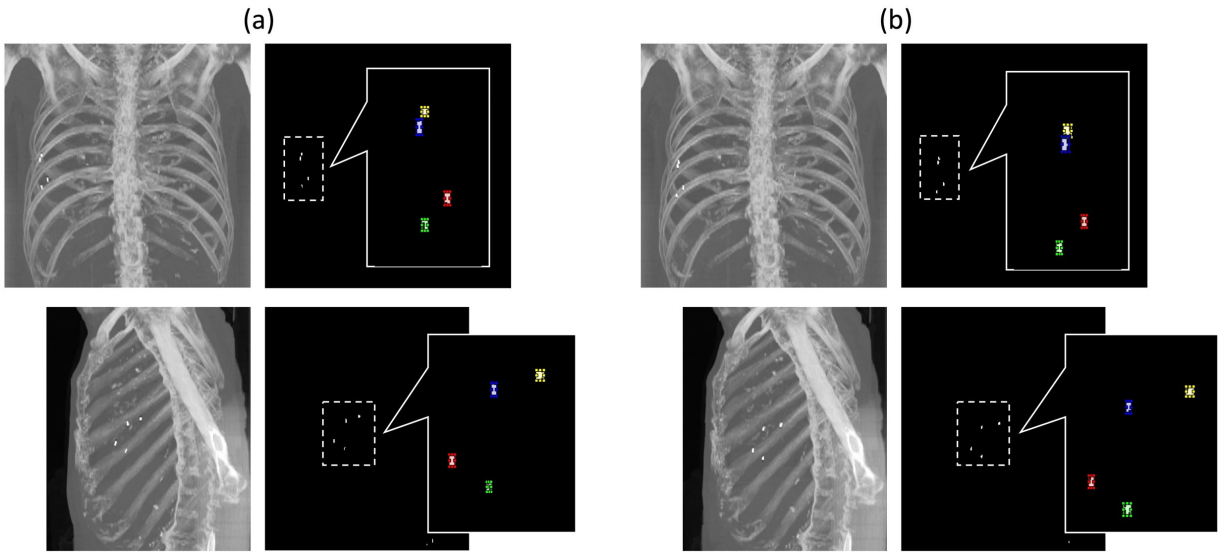


Fig. 5 (a) Maximum intensity projection image and binary image for marker segmentation generated in the anterior-posterior (top row) and left-right (bottom row) directions in phase 50. The colored rectangles represent the marker enclosing area and its center was used to define the three-dimensional marker position. Images generated in phase 0 are shown in (b). The motion of the fiducial markers between phase 0 and 50 due to respiration were confirmed.

Examples of the PVE distribution for each deformation in the LR, AP, and SI directions evaluated in patient #1 are shown in Figure 6. Relatively high PVEs were obtained around the tumor and the fiducial markers. The median, 25th, and 75th percentile of PVEs for each patient are summarized in Table 2. Focusing on the median values, high PVEs of more than 0.8 for deformations in all directions were obtained in most cases. These results suggest that there was reasonable linearity between the displacement of fiducial markers and the internal deformation in the evaluation region.

Table 2. Median (25th - 75th percentile) of percent variance explained (PVE) representing the strength of relationship between the displacement of the markers and the deformation for each patient.

	#1	#2	#3	#4	#5	#6
LR	0.82 (0.60 – 0.93)	0.88 (0.61 – 0.97)	0.86 (0.66 – 95)	0.78 (0.49 – 0.92)	0.80 (0.58 – 0.90)	0.80 (0.53 – 0.94)
AP	0.88 (0.64 – 0.96)	0.97 (0.86 – 0.99)	0.88 (0.72 – 95)	0.87 (0.63 – 0.94)	0.76 (0.49 – 0.90)	0.91 (0.65 – 0.97)
SI	0.80 (0.58 – 0.92)	0.95 (0.77 – 0.99)	0.84 (0.50 – 0.95)	0.84 (0.53 – 0.94)	0.87 (0.65 – 0.95)	0.78 (0.52 – 0.94)

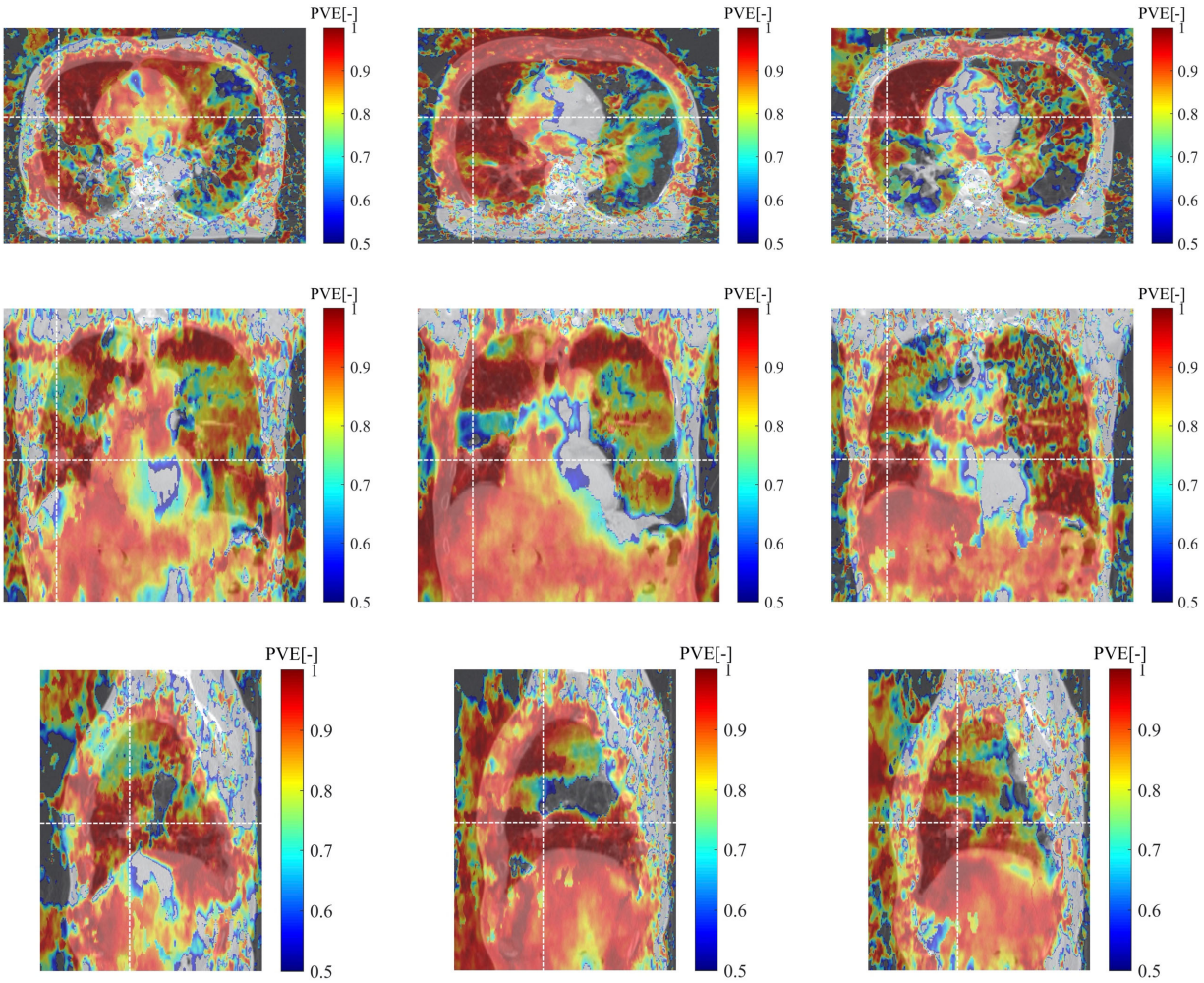
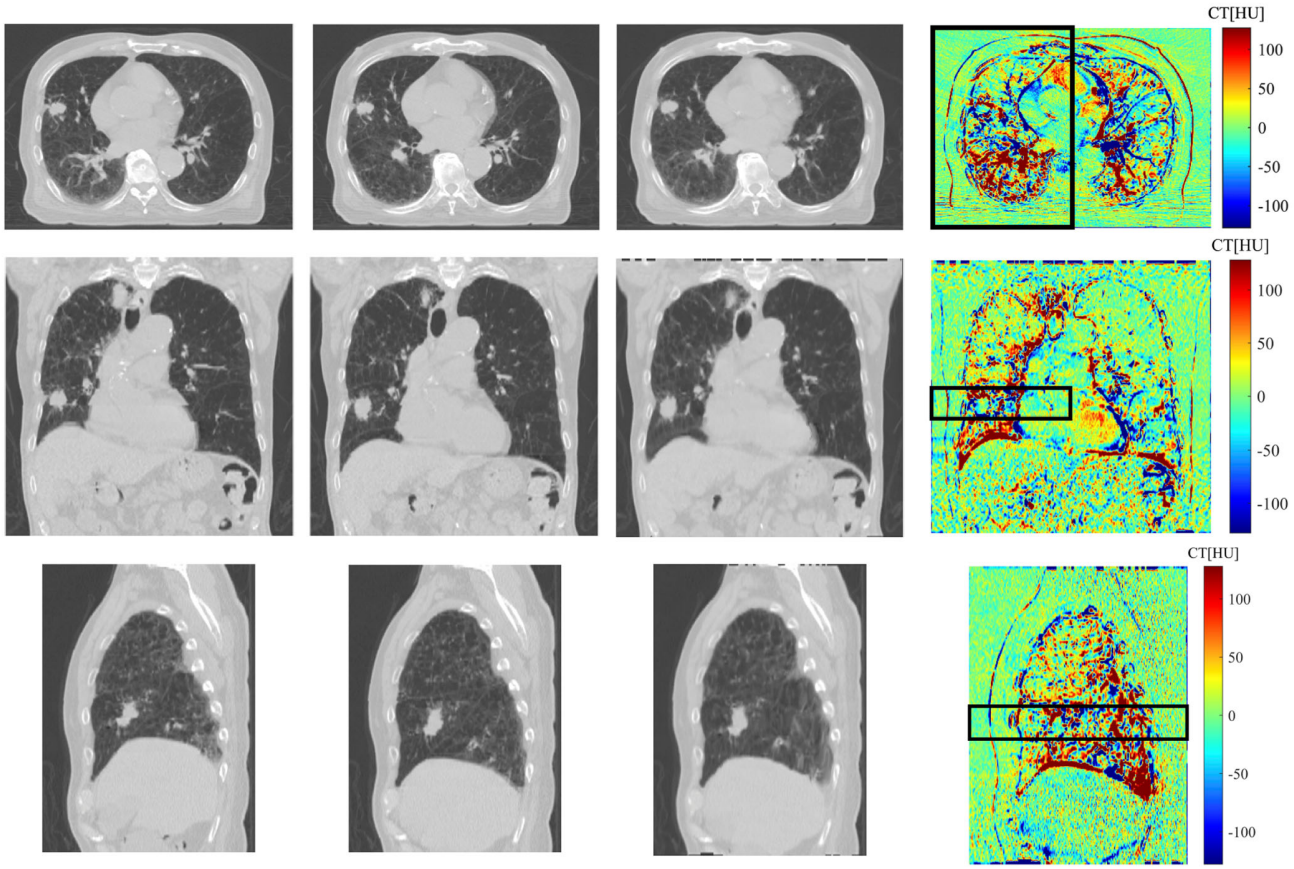


Fig. 6 The distribution of percent variance explained (PVE) obtained in the partial least squares regression (PLSR) between the displacement of the fiducial markers and the internal deformation in the left-right (left column), anterior-posterior (center column), and superior-inferior (right column) directions. Axial, coronal, and sagittal planes including the target center are shown in the top, middle, and bottom rows.

263 Examples of reference, ground truth, synthesized, and difference images for phase 0 in patient #1 are shown in
264 Figure 7. The internal anatomical structure, including the tumor shape, was well reconstructed in the synthesized
265 image. Since cardiac motion cannot be modeled by the proposed method, differences in CT values around the heart
266 were increased. In this example, differences in CT values in the back side region tended to increase since the
267 linearity in phase 0 was probably decreased. Note that CT dataset of phase 0 was not included in the training dataset.
268 To quantify the accuracy of the image synthesis, box and whisker plots of the absolute differences in CT values,
269 RMSE, NRMSE and SSIM, were evaluated. Differences in CT values may be small in regions where deformations
270 were relatively small, such as the upper lung region. Hence, to avoid underestimations, evaluation areas were limited
271 by three-dimensional regions, including the body half represented by a black solid line shown in the difference
272 images in Figure 7.

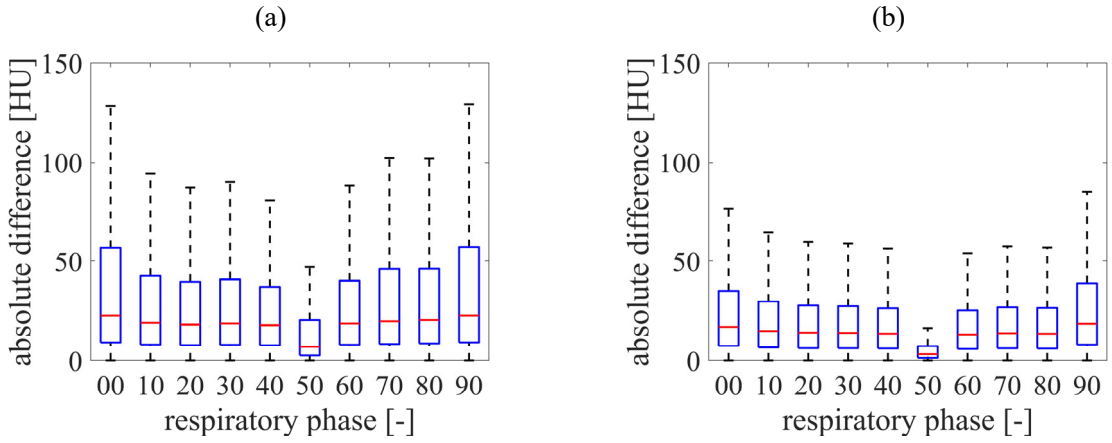
273



274

Fig. 7 From left column, reference image of phase 50, ground truth image of phase 0, synthesized image targeting phase 0 and difference image between the ground truth image and synthesized image. Axial, coronal and sagittal plane are shown from top row. Black solid rectangular in the difference image represent the evaluation region of PVE, RMSE, NRMSE and SSIM.

To show the distribution of the absolute differences in CT values between the ground truth image and the synthesized image, box plots for each patient and each respiratory phase are shown in Figure 8. The bottoms and tops of the boxes represent the 25% point (Q1) and 75% point (Q3) of the absolute difference in the CT values, respectively. The lower and upper whiskers are Q1–1.5 × (Q3–Q1) and Q3+1.5 × (Q3–Q1). Large differences were observed around the fiducial markers and bones having high CT values. The ratio of voxels with a large discrepancy (more than upper whisker) to all voxels in the evaluation area was about 10%. The samples having large differences were regarded as outliers and not shown in the plot. In all patients, absolute differences tended to increase in the inhale respiratory phase and the distribution expanded gradually. The median was less than 25 HU in most cases.



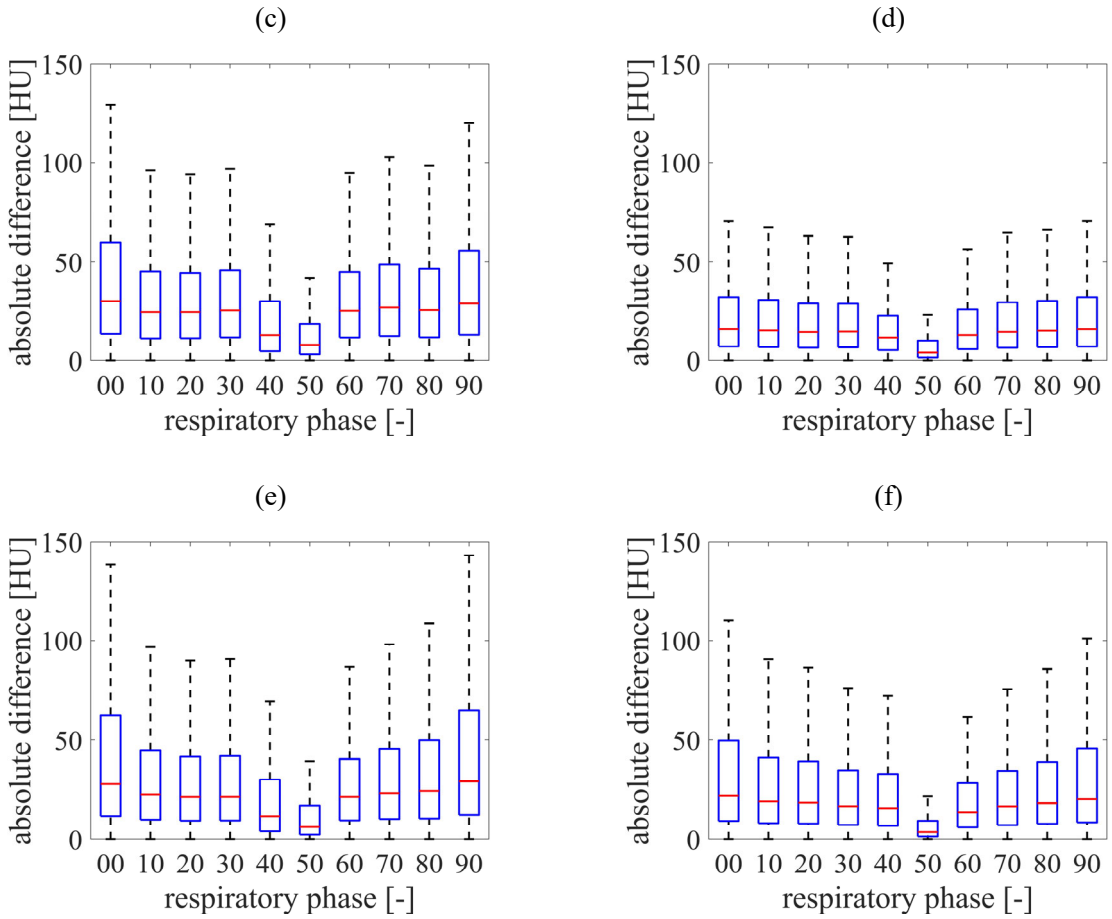


Fig. 8 Boxplots of the absolute difference of CT value between the ground truth image and the synthesized image for each patient. Listed from patient #1 to #6 in order from (a) to (f).

To quantify the accuracy of image synthesis, RMSE and NRMSE are summarized in Tables 3 and 4, respectively. For comparison to the benchmark, RMSE and NRMSE evaluated with the registered image by DIR are also summarized. As shown in Table 3, RMSEs from phase 10 to phase 80, corresponding to validation with the training dataset, were comparable to the benchmark performance in all patients. Since the linear regression by PLSR can addresses multicollinearity in multiple explanatory variables, the volumetric images synthesized from the input within the range of the training dataset were expected to be stable. In the evaluation of phases 0 and 90, which had marker displacement larger than that of the training dataset, the mean RMSEs were about 100 HU. The trends in NRMSEs were similar to the trends in RMSEs. Hence, there was less dependence of the image synthesis accuracy

299 on the distribution of CT values. SSIM are summarized in Table 5. As with the evaluations of RMSE and NRMSE,
300 SSIMs evaluated from phase 10 to phase 80 were comparable to the benchmark performance in all patients, although
301 SSIMs in phase 0 and 90 were decreased a little. These results suggested that the both image intensity and structure
302 of the synthesized images could be comparative those of DIR, although the accuracy could be reduced when the
303 marker displacement exceeds the range of training dataset.

304 The processing time to generate the volumetric images using the proposed method was less than 100 ms using a
305 conventional CPU (Intel(R) i9-9820X 3.30GHz). Thus, the proposed technique can be applied to real-time processes
306 during treatment.

307 **Table 3.** The root mean squared error (RMSE) [HU] of image synthesis for each patient. For reference, RMSEs of
308 deformable image registration (DIR) are shown at the bottom of each cell.

phase	#1	#2	#3	#4	#5	#6	Mean ± SD
0	111.1	85.4	135.7	62.8	110.0	121.5	104.4 ± 22.2
	80.2	68.8	83.3	57.1	63.5	76.8	71.6 ± 8.6
10	77.9	70.3	75.4	62.5	63.4	78.3	71.3 ± 6.0
	75.0	69.1	73.2	59.7	55.8	78.1	68.5 ± 7.5
20	68.1	68.5	69.9	62.4	58.6	74.3	67.0 ± 7.5
	60.7	66.1	63.5	65.0	55.3	66.0	62.8 ± 4.8
30	72.5	64.9	67.1	61.7	63.9	81.0	68.5 ± 6.0
	57.4	63.1	61.4	61.0	48.0	76.1	61.2 ± 7.7
40	60.4	63.1	58.6	59.7	49.0	72.3	60.5 ± 6.4
	49.1	60.8	54.1	53.8	31.8	70.4	53.3 ± 10.9
50	50.3	57.1	61.4	56.0	41.5	48.9	52.5 ± 6.0
	-	-	-	-	-	-	-
60	72.9	65.1	65.4	65.5	48.4	69.5	64.5 ± 7.1
	55.8	63.9	58.7	63.0	40.9	58.2	56.8 ± 7.1
70	100.5	67.3	71.4	65.1	65.1	74.5	74.0 ± 11.4
	101.3	64.6	67.0	65.0	54.0	69.9	70.3 ± 13.6
80	89.4	66.4	72.8	61.4	68.7	87.2	74.3 ± 9.7
	82.1	65.3	69.4	60.9	64.5	78.1	70.0 ± 7.1
90	113.2	95.0	116.6	63.2	105.5	106.2	100.0 ±1 6.5
	81.4	67.7	77.9	60.3	66.2	91.6	74.2 ± 9.8

Mean	81.6 ± 20.2	70.3 ± 10.7	79.4 ± 24.2	62.0 ± 2.6	67.4 ± 21.8	81.4 ± 19.1	73.7±19.6
± SD	64.3 ± 26.2	58.9 ± 19.8	60.8 ± 22.0	54.6 ± 18.5	48.0 ± 19.0	66.5 ± 23.7	58.9 ± 22.6

Table 4. The normalized root mean squared error (NRMSE) [%] of image synthesis for each patient. For reference, NRMSEs of deformable image registration (DIR) are shown at the bottom of each cell.

phase	#1	#2	#3	#4	#5	#6	Mean ± SD
0	15.2	11.3	15.2	7.6	14.6	15.2	13.2 ± 2.7
	11.0	9.1	9.4	6.9	8.4	9.6	9.1 ± 1.1
10	10.6	9.3	8.5	7.5	8.2	9.8	9.0 ± 0.9
	10.2	9.1	8.2	7.2	7.2	9.8	8.6 ± 1.1
20	9.2	9.0	7.8	7.5	7.5	9.2	8.4 ± 0.7
	8.2	8.7	7.1	7.8	7.1	8.2	7.8 ± 0.5
30	9.8	8.5	7.5	7.4	8.1	9.9	8.5 ± 0.9
	7.8	8.3	6.9	7.3	6.1	9.3	7.6 ± 0.9
40	8.2	8.3	6.5	7.1	6.1	8.8	7.5 ± 0.9
	6.6	8.0	6.0	6.4	4.0	8.6	6.6 ± 1.4
50	6.8	7.5	6.8	6.7	5.2	6.0	6.5 ± 0.7
	-	-	-	-	-	-	-
60	9.7	8.5	7.3	7.8	6.0	8.5	8.0 ± 1.1
	7.5	8.3	6.5	7.5	5.1	7.2	7.0 ± 0.9
70	13.4	8.8	7.9	7.8	8.2	9.1	9.2 ± 1.8
	13.5	8.4	7.5	7.8	6.8	8.6	8.8 ± 2.1
80	12.0	8.7	8.1	7.4	9.0	10.7	9.3 ± 1.4
	11.0	8.5	7.7	7.3	8.5	9.5	8.8 ± 1.1
90	15.5	12.6	13.1	7.6	13.9	13.0	12.6 ± 2.2
	11.1	9.0	8.7	7.2	8.7	11.2	9.3 ± 1.3
Mean ±	11.0 ± 2.8	9.2 ± 1.4	8.9 ± 2.7	7.4 ± 0.3	8.7 ± 3.0	10.0 ± 2.4	9.2 ± 2.6
SD	8.7 ± 3.5	7.7 ± 2.6	6.8 ± 2.5	6.5 ± 2.2	6.2 ± 2.5	8.2 ± 2.9	7.4 ± 2.9

Table 5. The structural similarity index measure (SSIM) for each patient. For reference, SSIMs of deformable image registration (DIR) are shown at the bottom of each cell.

phase	#1	#2	#3	#4	#5	#6	Mean ± SD
0	0.78	0.88	0.70	0.93	0.79	0.81	0.82 ± 0.08
	0.85	0.92	0.82	0.94	0.88	0.88	0.88 ± 0.04
10	0.85	0.92	0.82	0.93	0.88	0.88	0.88 ± 0.04
	0.85	0.92	0.83	0.93	0.90	0.88	0.89 ± 0.04

20	0.87	0.92	0.83	0.93	0.89	0.89	0.89 ± 0.04
	0.88	0.93	0.85	0.94	0.90	0.90	0.90 ± 0.03
30	0.86	0.93	0.82	0.93	0.89	0.89	0.89 ± 0.04
	0.89	0.93	0.84	0.94	0.91	0.91	0.90 ± 0.03
40	0.86	0.93	0.82	0.93	0.89	0.89	0.89 ± 0.04
	0.89	0.93	0.84	0.94	0.91	0.91	0.90 ± 0.03
50	0.93	0.98	0.93	0.97	0.96	0.98	0.96 ± 0.02
	-	-	-	-	-	-	-
60	0.86	0.93	0.83	0.94	0.90	0.93	0.90 ± 0.05
	0.89	0.94	0.84	0.95	0.92	0.95	0.91 ± 0.04
70	0.81	0.93	0.81	0.93	0.88	0.90	0.88 ± 0.05
	0.83	0.93	0.83	0.93	0.90	0.91	0.89 ± 0.05
80	0.82	0.93	0.82	0.93	0.87	0.87	0.87 ± 0.05
	0.83	0.93	0.83	0.93	0.88	0.90	0.88 ± 0.05
90	0.78	0.86	0.73	0.92	0.80	0.82	0.82 ± 0.07
	0.85	0.92	0.82	0.93	0.88	0.86	0.88 ± 0.04
Mean	0.84 ± 0.05	0.92 ±0.03	0.82 ±0.07	0.94 ± 0.01	0.88 ± 0.05	0.89 ± 0.05	0.88 ± 0.04
± SD	0.87 ± 0.03	0.93 ±0.01	0.84 ± 0.04	0.94 ± 0.01	0.90 ± 0.02	0.90 ± 0.02	0.90 ± 0.02

Accuracy of tumor localization in patient dataset

An example of tumor image in phase 0 of patient #1 is shown in Figure 9. The tumor shape was well reconstructed in the synthesized image. The accuracy of tumor localization and Dice coefficient for each patient are summarized in Table 6. In all evaluated patients, three-dimensional tumor localization error was less than 1.0 mm from phase 10 to 80 with Dice coefficient more than about 0.9, which suggested that the tumor shape was reasonably reproduced. Although patient #3 exhibited a large tumor localization error, the three-dimensional tumor localization error in phase 0 and 90 of the other patients was less than 2.0 mm.

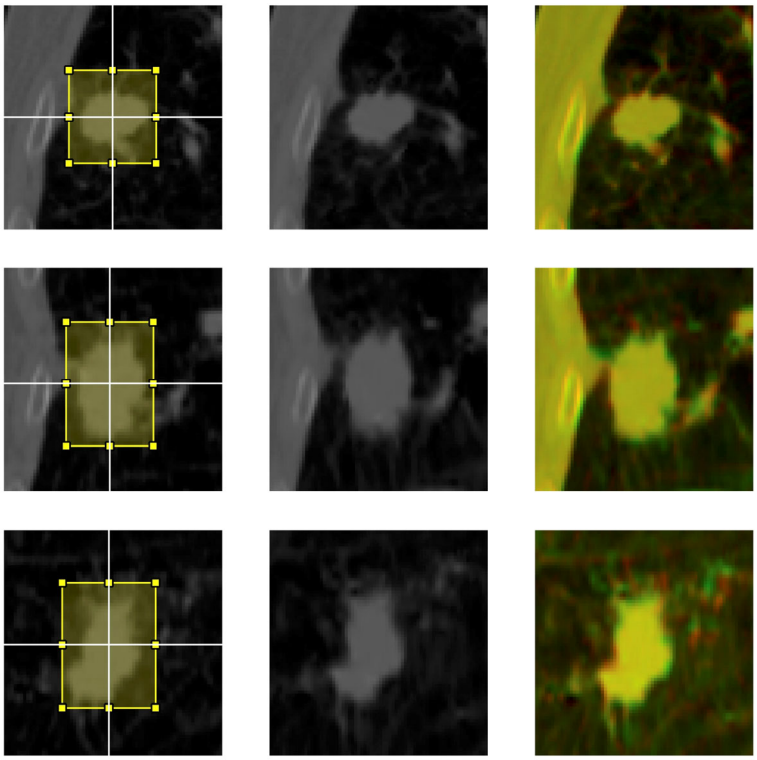


Fig. 9 Left column: The ground truth images around the tumor in phase 0 of patient #1. The yellow rectangular region represents the tumor region. Center column: The synthesized image cropped at the same position as the ground truth image. Right column: The overlay of the ground truth and synthesized images. The green and red regions indicate the mismatch of image intensity. The axial, coronal, and sagittal planes are shown from top to bottom.

Table 6. Tumor localization error [mm] in the left-right (LR), superior-inferior (SI) and anterior-posterior (AP) direction. Dice coefficient was also summarized as a metric of similarity of tumor shape.

Phase		#1	#3	#5	#6	Mean ± SD
0	LR	0.5	-0.3	0.8	-1.2	0.0 ± 0.9
	SI	1.5	2.4	0.8	-0.9	1.0 ± 1.4
	AP	-0.5	-2.7	0.5	1.1	-0.4 ± 1.7
	3D	1.7	3.6	1.2	1.9	2.1 ± 1.1
	Dice	0.85	0.73	0.78	0.88	0.81 ± 0.07
10	LR	0.0	-0.5	0.1	-0.7	-0.3 ± 0.4
	SI	-0.6	-0.5	0.6	-0.7	-0.3 ± 0.6
	AP	-0.2	0.3	0.2	0.6	0.2 ± 0.3
	3D	0.6	0.7	0.6	1.2	0.8 ± 0.3
	Dice	0.91	0.91	0.83	0.91	0.89 ± 0.04
20	LR	-0.1	0.0	-0.1	-0.5	-0.2 ± 0.2
	SI	0.0	0.7	0.5	-1.2	0.0 ± 0.8
	AP	0.5	0.2	0.0	0.4	0.3 ± 0.2

	3D	0.5	0.7	0.5	1.4	0.8 ± 0.4
	Dice	0.91	0.93	0.86	0.88	0.90 ± 0.03
30	LR	0.0	-1.0	-0.1	-0.2	-0.3 ± 0.5
	SI	0.0	0.5	-0.1	-0.8	-0.1 ± 0.5
	AP	0.3	0.0	0.6	0.1	0.2 ± 0.3
	3D	0.3	1.1	0.6	0.8	0.7 ± 0.3
	Dice	0.90	0.91	0.86	0.91	0.89 ± 0.02
40	LR	-0.1	-0.1	-0.1	-0.1	-0.1 ± 0.0
	SI	-0.1	-0.4	0.0	-0.6	-0.3 ± 0.3
	AP	0.1	-0.2	0.5	0.2	0.1 ± 0.3
	3D	0.2	0.4	0.5	0.6	0.4 ± 0.2
	Dice	0.90	0.95	0.92	0.89	0.92 ± 0.03
50	LR	0.1	-0.1	0.0	0.0	0.0 ± 0.1
	SI	0.1	0.5	0.2	-0.2	0.1 ± 0.3
	AP	0.0	0.4	0.5	0.1	0.3 ± 0.2
	3D	0.1	0.6	0.6	0.2	0.4 ± 0.3
	Dice	0.94	0.95	0.93	0.97	0.95 ± 0.02
60	LR	-0.2	-0.2	0.1	0.2	0.0 ± 0.2
	SI	0.0	-0.1	0.0	-0.2	-0.1 ± 0.1
	AP	0.0	-0.1	0.3	0.0	0.1 ± 0.2
	3D	0.2	0.2	0.4	0.3	0.3 ± 0.1
	Dice	0.90	0.95	0.91	0.94	0.93 ± 0.03
70	LR	0.2	0.2	0.3	-0.2	0.1 ± 0.2
	SI	-0.5	-0.1	0.8	0.0	0.1 ± 0.5
	AP	0.3	0.0	-0.1	0.4	0.1 ± 0.3
	3D	0.6	0.2	0.9	0.5	0.5 ± 0.3
	Dice	0.86	0.93	0.85	0.93	0.89 ± 0.04
80	LR	0.0	0.0	0.1	0.5	0.2 ± 0.2
	SI	0.3	0.1	0.9	0.0	0.3 ± 0.4
	AP	0.1	-0.2	-0.2	-0.3	-0.2 ± 0.2
	3D	0.3	0.3	1.0	0.6	0.5 ± 0.3
	Dice	0.89	0.94	0.83	0.92	0.90 ± 0.05
90	LR	0.6	0.3	0.5	0.3	0.4 ± 0.1
	SI	1.7	3.8	0.5	-0.6	1.4 ± 1.9
	AP	-0.6	-1.8	0.0	-0.3	-0.7 ± 0.8
	3D	1.9	4.2	0.7	0.7	1.9 ± 1.7
	Dice	0.84	0.76	0.81	0.89	0.83 ± 0.06
Mean ± SD	LR	0.1 ± 0.3	-0.1 ± 0.4	0.2 ± 0.3	-0.2 ± 0.5	0.0 ± 0.2
	SI	0.2 ± 0.8	0.7 ± 1.4	0.4 ± 0.4	−0.5 ± 0.4	0.2 ± 0.5
	AP	0.0 ± 0.3	-0.4 ± 1.0	0.2 ± 0.4	0.2 ± 0.4	0.0 ± 0.3
	3D	0.6 ± 0.6	1.2 ± 1.5	0.9 ± 0.4	0.8 ± 0.5	0.8 ± 0.6
	Dice	0.89 ± 0.03	0.90 ± 0.08	0.86 ± 0.05	0.91 ± 0.03	0.89 ± 0.04

Discussion

We presented a detailed evaluation of image synthesis accuracy and tumor localization accuracy for the proposed real-time CT image generation technique. The accuracies are similar to previous studies. Mishra et al. proposed volumetric imaging based on principal component analysis (PCA).[25] They evaluated NRMSE, assuming the variation of motion magnitude of the chest wall and diaphragm, using XCAT phantom. In case of large variations of motion amplitude, NRMSE increased up to about 15%. Li et al. also demonstrated real-time volumetric imaging based on PCA[24] and the mean NRMSE was 6.9 ± 2.4% when the respiratory motion amplitude changed between the 4DCT acquisition and the treatment session. NRMSEs evaluated in this study were 9.2±2.6% and 7.5±1.1% in patient 4DCT dataset and XCAT phantom, respectively. Although the testing protocol, data for validation, and algorithm of DIR were different, the imaging accuracy of the proposed technique is expected to be similar to those of previous research.

The three-dimensional tumor localization accuracy in ten respiratory phases for the patient 4DCT dataset was 0.8 ± 0.6 mm and 0.4 ± 0.3 mm for the XCAT phantom in this study. These were equivalent to those of previous studies, including the study of Li et al.[19], which demonstrated a three-dimensional accuracy of 0.8 ± 0.5 mm, and the study of Mishra et al.[20], which demonstrated an accuracy of 1.0 ± 0.9 mm. The accuracy in phases 0 and 90 of the patient 4DCT dataset was less than 2 mm, except for patient #3, and is considered clinically acceptable. Three-dimensional tumor localization accuracy in phases 0 and 90 of patient #3 was about 4 mm. In patient #3, variations in the difference of motion range were relatively large compared with the other cases. Linearity around the tumor may have been decreased in phases 0 and 90 due to the large motion variation. A more robust technique like deep learning, which can address nonlinearity, may be applicable.

Motion artifact in 4DCT is a major source of error since the proposed volumetric imaging utilizes the CT datasets to derive marker displacement and internal deformation. CT reconstruction algorithms, which can reduce the artifacts, will be more effective[27,28] in obtaining accurate deformations. Accuracy of image synthesis may be decreased in regions far from the fiducial markers since linearity between marker displacement and the internal deformation may decrease. Inferring models, such as deep learning, can be combined with the proposed model to address nonlinearity. In this study, patient setup error or inter-fractional anatomical variations, such as tumor

363 shrinkage or weight change, were not simulated in the evaluation, since a series of 4DCT dataset was used for
364 testing. As for initial patient setup, the current image guidance techniques, such as CBCT, can correct patient
365 position with an accuracy of less than 1 mm.[29,30] However, the current volumetric imaging techniques, including
366 the previous studies, could not handle large internal anatomical variations between 4DCT acquisition and treatment
367 sessions. These variations may cause large errors in image synthesis and tumor localization. To adapt to these
368 variations, 4D-CBCT, acquired just before the treatment, could be utilized for updating the evaluation model.

369 In this study, we assumed an orthogonal x-ray imaging system to obtain the three-dimensional position of the
370 internal fiducial markers. The fiducial markers such as gold sphere of which diameter is 1.5 or 2.0 mm can be
371 recognized in x-ray images by means of pattern matching technique in most cases in lung or liver. Note that there
372 could be a risk of miss detection of the markers in one or both x-ray images if the fiducial markers such as thin
373 coiled shape are used[31]. In the current RTRT, the fiducial markers should be recognized in both x-ray images for
374 three-dimensional calculation. On the other hand, the treatment systems that have dual imaging functions during
375 treatment are not widely used yet. Li et al. proposed deriving the three-dimensional position from a single planar x-
376 ray image.[32] A study on volumetric imaging utilizing the fiducial markers with a single x-ray imaging device, which
377 can be widely applied in most clinical systems, is the next step of investigation for the proposed technique.

378 Regarding clinical significance, the proposed CT image generation technique can be applied to real-time beam
379 gating/tracking as a motion management technique based on tumor position and anatomical structure of the
380 surrounding organs, especially for lungs in SBRT and particle therapy. For instance, since the volumetric images
381 have CT values similar to treatment planning CTs, treatment beam gating based on the evaluation of water
382 equivalent thickness for particle therapy[33] could be realized. Synthesized volumetric images could be applied for
383 both real-time dose calculations during treatment[34] and retrospective dose accumulations. In addition, imaging doses
384 due to continuous x-ray imaging can be minimized by collimating the imaging region using only the marker area.
385 In all volumetric imaging techniques, the imaging accuracy will be decreased when actual motion is different from
386 the training dataset. Displacement of the fiducial markers could also be used for interlock-function. For instance,
387 when the marker displacement exceeds the threshold value, the treatment beam can be forced off. The volumetric

388 imaging technique using the fiducial markers can be one of the options due to its effectiveness, although the
389 technique requires marker insertion and an additional imaging dose during treatment.

390 Conclusion

391 In this study, we proposed a real-time CT image generation technique utilizing the positional displacement of
392 internal fiducial markers evaluated in real-time during the treatment. The XCAT phantom dataset and patient 4DCT
393 dataset, including ten CT datasets for each respiratory phase, were used for the evaluation of image synthesis
394 accuracy and tumor localization accuracy. In the validation with XCAT phantom, the mean NRMSE, SSIM and
395 three-dimensional tumor localization error were 7.5±1.1%, 0.95±0.02 and 0.4±0.3 mm, respectively. In the
396 evaluation with patient 4DCT dataset, eight CT datasets, excluding two CT datasets of the inhale respiratory phase,
397 were used in the training process. CT images were generated for each of the ten respiratory phases using the
398 corresponding marker displacements and compared with the original CT dataset. RMSEs, NRMSEs and SSIMS
399 from phase 10 to 80, corresponding to the training dataset, were comparable to those of DIR. The three-dimensional
400 tumor localization error was less than 1 mm. In the evaluation of phases 0 and 90, which had marker displacement
401 larger than that of the training dataset, the mean RMSEs and NRMSEs were about 100 HU and 13%, respectively,
402 and were comparable to previous studies. Three-dimensional tumor localization errors in phases 0 and 90 were less
403 than 2.0 mm, except for one case having large motion variation. These results demonstrate the feasibility of the
404 proposed real-time CT image generation technique.

References

1. Furukawa T, Inaniwa T, Sato S, et al. Moving target irradiation with fast rescanning and gating in particle therapy [published online ahead of print 2010/10/23]. *Med Phys.* 2010;37(9):4874-4879.
2. Ohara K, Okumura T, Akisada M, et al. Irradiation synchronized with respiration gate [published online ahead of print 1989/10/01]. *Int J Radiat Oncol Biol Phys.* 1989;17(4):853-857.
3. Minohara S, Kanai T, Endo M, Noda K, Kanazawa M. Respiratory gated irradiation system for heavy-ion radiotherapy [published online ahead of print 2000/06/23]. *Int J Radiat Oncol Biol Phys.* 2000;47(4):1097-1103.
4. Bert C, Saito N, Schmidt A, Chaudhri N, Schardt D, Rietzel E. Target motion tracking with a scanned particle beam [published online ahead of print 2008/01/17]. *Med Phys.* 2007;34(12):4768-4771.
5. Park YK, Son TG, Kim H, et al. Development of real-time motion verification system using in-room optical images for respiratory-gated radiotherapy [published online ahead of print 2013/09/17]. *Journal of applied clinical medical physics / American College of Medical Physics.* 2013;14(5):25-42.
6. Shirato H, Oita M, Fujita K, Watanabe Y, Miyasaka K. Feasibility of synchronization of real-time tumor-tracking radiotherapy and intensity-modulated radiotherapy from viewpoint of excessive dose from fluoroscopy [published online ahead of print 2004/09/01]. *Int J Radiat Oncol Biol Phys.* 2004;60(1):335-341.
7. Inoue T, Katoh N, Onimaru R, et al. Stereotactic body radiotherapy using gated radiotherapy with real-time tumor-tracking for stage I non-small cell lung cancer. *Radiat Oncol.* 2013;8:69.
8. Katoh N, Soda I, Tamamura H, et al. Clinical outcomes of stage I and IIA non-small cell lung cancer patients treated with stereotactic body radiotherapy using a real-time tumor-tracking radiotherapy system. *Radiat Oncol.* 2017;12(1):3.
9. Uchinami Y, Katoh N, Abo D, et al. Treatment outcomes of stereotactic body radiation therapy using a real-time tumor-tracking radiotherapy system for hepatocellular carcinomas [published online ahead of print 2021/04/25]. *Hepatol Res.* 2021. doi: 10.1111/hepr.13649.
10. Shimizu S, Miyamoto N, Matsuura T, et al. A proton beam therapy system dedicated to spot-scanning increases accuracy with moving tumors by real-time imaging and gating and reduces equipment size. *PLoS One.* 2014;9(4):e94971.
11. Matsuura T, Miyamoto N, Shimizu S, et al. Integration of a real-time tumor monitoring system into gated proton spot-scanning beam therapy: an initial phantom study using patient tumor trajectory data [published online ahead of print 2013/07/05]. *Med Phys.* 2013;40(7):071729.
12. Keall PJ, Colvill E, O'Brien R, et al. The first clinical implementation of electromagnetic transponder-guided MLC tracking. *Med Phys.* 2014;41(2):020702.

13. Seppenwoolde Y, Berbeco RI, Nishioka S, Shirato H, Heijmen B. Accuracy of tumor motion compensation algorithm from a robotic respiratory tracking system: a simulation study. *Med Phys.* 2007;34(7):2774-2784.
14. Depuydt T, Verellen D, Haas O, et al. Geometric accuracy of a novel gimbals based radiation therapy tumor tracking system. *Radiother Oncol.* 2011;98(3):365-372.
15. Li R, Lewis JH, Jia X, et al. On a PCA-based lung motion model [published online ahead of print 2011/08/26]. *Phys Med Biol.* 2011;56(18):6009-6030.
16. Zhang Q, Pevsner A, Hertanto A, et al. A patient-specific respiratory model of anatomical motion for radiation treatment planning [published online ahead of print 2008/01/17]. *Med Phys.* 2007;34(12):4772-4781.
17. Wei R, Zhou F, Liu B, et al. Convolutional Neural Network (CNN) Based Three Dimensional Tumor Localization Using Single X-Ray Projection. *IEEE Access.* 2019;7:37026-37038.
18. Wei R, Zhou F, Liu B, et al. Real-time tumor localization with single x-ray projection at arbitrary gantry angles using a convolutional neural network (CNN) [published online ahead of print 2020/01/03]. *Phys Med Biol.* 2020;65(6):065012.
19. Lei Y, Tian Z, Wang T, et al. Deep learning-based real-time volumetric imaging for lung stereotactic body radiation therapy: a proof of concept study [published online ahead of print 2020/10/21]. *Phys Med Biol.* 2020;65(23):235003.
20. Miyamoto N, Ishikawa M, Bengua G, et al. Optimization of fluoroscopy parameters using pattern matching prediction in the real-time tumor-tracking radiotherapy system. *Phys Med Biol.* 2011;56(15):4803-4813.
21. Segars WP, Sturgeon G, Mendonca S, Grimes J, Tsui BM. 4D XCAT phantom for multimodality imaging research [published online ahead of print 2010/10/23]. *Med Phys.* 2010;37(9):4902-4915.
22. Wold S, Sjostrom M, Eriksson L. PLS-regression: a basic tool of chemometrics. *Chemometr Intell Lab.* 2001;58(2):109-130.
23. de Jong S. SIMPLS: An alternative approach to partial least squares regression. *Chemometr Intell Lab.* 1993;18(3):251-263.
24. Li R, Jia X, Lewis JH, et al. Real-time volumetric image reconstruction and 3D tumor localization based on a single x-ray projection image for lung cancer radiotherapy [published online ahead of print 2010/07/17]. *Med Phys.* 2010;37(6):2822-2826.
25. Mishra P, Li R, James SS, et al. Evaluation of 3D fluoroscopic image generation from a single planar treatment image on patient data with a modified XCAT phantom [published online ahead of print 2013/01/23]. *Phys Med Biol.* 2013;58(4):841-858.
26. Harada K, Katoh N, Suzuki R, et al. Evaluation of the motion of lung tumors during stereotactic body radiation therapy (SBRT) with four-dimensional computed tomography (4DCT) using real-time tumor-tracking radiotherapy system (RTRT). *Phys Med.* 2016. doi: 10.1016/j.ejmp.2015.10.093.
27. Gianoli C, Riboldi M, Spadea MF, et al. A multiple points method for 4D CT image sorting [published online ahead of print 2011/04/02]. *Med Phys.* 2011;38(2):656-667.
28. Spadea MF, Baroni G, Gierga DP, Turcotte JC, Chen GT, Sharp GC. Evaluation and commissioning of a surface based system for respiratory sensing in 4D CT [published online ahead of print 2011/02/19]. *J Appl Clin Med Phys.* 2010;12(1):3288.
29. Munbodh R, Knisely JP, Jaffray DA, Moseley DJ. 2D-3D registration for cranial radiation therapy using a 3D kV CBCT and a single limited field-of-view 2D kV radiograph [published online ahead of print 2018/02/23]. *Med Phys.* 2018;45(5):1794-1810.

30. Liu RJ, Yang SX, Neylon J, et al. Residual setup errors in cranial stereotactic radiosurgery without six degree of freedom robotic couch: Frameless versus rigid immobilization systems [published online ahead of print 2020/02/19]. *J Appl Clin Med Phys.* 2020;21(3):87-93.
31. Miyamoto N, Maeda K, Abo D, et al. Quantitative evaluation of image recognition performance of fiducial markers in real-time tumor-tracking radiation therapy [published online ahead of print 2019/08/21]. *Phys Med.* 2019;65:33-39.
32. Li R, Fahimian BP, Xing L. A Bayesian approach to real-time 3D tumor localization via monoscopic x-ray imaging during treatment delivery [published online ahead of print 2011/08/24]. *Med Phys.* 2011;38(7):4205-4214.
33. Fassi A, Seregni M, Riboldi M, et al. Surrogate-driven deformable motion model for organ motion tracking in particle radiation therapy. *Phys Med Biol.* 2015;60(4):1565-1582.
34. Ravkilde T, Skouboe S, Hansen R, Worm E, Poulsen PR. First online real-time evaluation of motion-induced 4D dose errors during radiotherapy delivery [published online ahead of print 2018/06/06]. *Med Phys.* 2018. doi: 10.1002/mp.13037.

Acknowledgments

This research was partially supported by AMED under Grant Number JP20he2302001 and KAKENHI 18K19887, 20H03612. The authors would like to express their gratitude to Dr. Seiko Nishioka and Mr. Masaki Kawahara of the Department of Radiology, NTT-East Sapporo Hospital, Japan, for sharing the dataset of 3D tumor motion and external surrogate motion.

Conflict of Interest Statement

The authors have no relevant conflicts of interest to disclose.