Physica Medica 59 (2019) 22–29

Contents lists available at ScienceDirect

Physica Medica

journal homepage: www.elsevier.com/locate/ejmp

Original paper

Real-time tumor tracking using fluoroscopic imaging with deep neural network analysis

Ryusuke Hirai[a,*], Yukinobu Sakata[a], Akiyuki Tanizawa[a], Shinichiro Mori[b]

[a] Corporate Research and Development Center, Toshiba Corporation, Kanagawa 212-8582, Japan
[b] Research Center for Charged Particle Therapy, National Institute of Radiological Sciences, Inage-ku, Chiba 263-8555, Japan

ARTICLE INFO

Keywords:
Markerless tumor tracking
Deep neural network
Particle beam therapy
Image guidance

ABSTRACT

Purpose: To improve respiratory gating accuracy and treatment throughput, we developed a fluoroscopic markerless tumor tracking algorithm based on a deep neural network (DNN).

Methods: In the learning stage, target positions were projected onto digitally reconstructed radiography (DRR) images from four-dimensional computed tomography (4DCT). DRR images were cropped into subimages of the target or surrounding regions to build a network that takes input of the image pattern of subimages and produces a target probability map (TPM) for estimating the target position. Using multiple subimages, a DNN was trained to generate a TPM based on the target position projected onto the DRRs. In the tracking stage, the network takes in the subimages cropped from fluoroscopic images at the same position of the subimages on the DRRs and produces TPMs, which are used to estimate target positions. We integrated the lateral correction to modify an estimated target position by using a linear regression model. We tracked five lung and five liver cases, and calculated tracking accuracy (Euclidian distance in 3D space) by subtracting the estimated position from the reference.

Results: Tracking accuracy averaged over all patients was 1.64 ± 0.73 mm. Accuracy for liver cases (1.37 ± 0.81 mm) was better than that for lung cases (1.90 ± 0.65 mm). Computation time was < 40 ms for a pair of fluoroscopic images.

Conclusions: Our markerless tracking algorithm successfully estimated tumor positions. We believe our results will provide useful information to advance tumor tracking technology.

1. Introduction

The advantages of carbon-ion beam treatment over proton and photon beam therapy is its improved dose conformation to the target and minimized exposure to OARs. Several researchers have aimed to increase the positional accuracy of carbon-ion beam treatment when tracking a target [1,2] Particularly in the thoracoabdominal region, which is subject to both intrafractional and interfractional motion, most treatment centers perform respiratory gating to guide treatment. Two major respiratory gating techniques have been developed: external surrogate motion tracking and internal tumor tracking. Kubo et al. first integrated the external gating technique into proton beam therapy using an abdominal surface motion sensor [3]. However, this technique is disadvantaged by the assumption that the observed external respiratory signal closely correlates with actual internal tumor motion [4].

In contrast, Shirato et al. (2003) developed fluoroscopic-based fiducial marker tracking as an internal tumor tracking technique [5]. This technique solves the problem of external gating by directly identifing marker position. However, we hypothesized that fluoroscopic-based fiducial marker tracking may be become inaccurate when the marker is not implanted in the tumor. One of the drawbacks includes pneumothorax [6]. To solve this, several methods that do not require fiducial markers (markerless tracking) are now in clinical use [7].

One approach to markerless tracking using imaging patterns is multiple-template matching using fluoroscopic images [8]. This method requires a template image preparation process before irradiation, however, which might slow treatment throughput [8]. To solve this, several methods have been proposed. Li et al. [9] suggested the preparation of template images using treatment planning 4DCT images. This method overcomes the differences in quality between digitally reconstructed radiography (DRR) and fluoroscopic images by using subtraction template images from temporally averaged images. Shieh et al. used a Bayesian approach to markerless lung tumor tracking using

* Corresponding author.
E-mail address: ryusuke.hirai@toshiba.co.jp (R. Hirai).

https://doi.org/10.1016/j.ejmp.2019.02.006
Received 26 September 2018; Received in revised form 29 January 2019; Accepted 9 February 2019
Available online 25 February 2019

4DCT data [22]. Their algorithm used template matching of the tumor features. Other methods aim to improve treatment throughput through the use of image pattern features which employ the mean-shift method or level set method [10,11]. However, these methods require calculation of tumor position on the first frame, which is difficult with images subject to motion artifact.

Another approach to markerless tracking uses a linear regression model with correlation information of the relationship of the tumor position to the position of the diaphragm edge [12,13]. Although these methods can be applied to abdominal treatment sites such as the liver, fluoroscopic images in these cases should contain the liver dome. Moreover, tumor position should be input on fluoroscopic images before irradiation. However, treatment throughput and treatment accuracy with this method remain subject to motion artifact. Thus, despite these various improvements, an optimum method to ensure positional accuracy in carbon-ion beam treatment when tracking a target has yet to be developed.

To solve these problems, we developed a deep neural network (DNN)-based markerless tumor tracking methodology using four-dimensional planning computed tomography (4DCT) data. Since the tumor position is generated from these data, and the DNN develops the regression model from image patterns in one or more regions of interest (ROIs), all computation is performed before irradiation. The advantage of this algorithm is that even when not recording tumor position, it still identifies the motion pattern.

Here, we evaluated the tumor tracking accuracy of this markerless system using five lung and five liver cases.

2. Materials and methods

2.1. Image acquisition

A retrospective analysis was performed on 10 fluoroscopic image data sets (five liver and five lung data sets), which were processed as part of our treatment workflow [7]. In treatment planning, 4DCT images were acquired under free-breathing conditions using a 320-slice CT (Aquilion One Vision®, Canon Medical Systems, Otawara, Japan). 4DCT images were subdivided into 10 phases (T00: peak inhalation). CT section thickness was 1.0 mm. A board-certified oncologist manually contoured the gross tumor volume (GTV) on the CT data at T50, and the contours at other respiratory phases were then automatically generated using B-Spline-based deformable image registration [14,15]. Reference tumor positions in the respective phases were defined as the center of mass of the GTV. Fluoroscopic image data were acquired during treatment delivery using a dynamic flat panel detector (PaxScan 3030®, Varian Medical Systems, Palo Alto CA, USA) installed on each side of the vertical irradiation port, with the corresponding x-ray tubes installed under the floor of our treatment room. The distance from the tube to the room isocenter and the detector were 170 cm and 240 cm, respectively. Each fluoroscopic data set consisted of 200 frames acquired at 15 frames per second (fps). Fluoroscopic image matrix size was 768 × 768, with a pixel size of 0.388 × 0.388 mm.

2.2. Tracking algorithm

Our tumor tracking algorithm consisted of 6 steps; steps 1–3 were learning stages acquired before a treatment and steps 4–6 were tracking stages acquired during beam irradiation (Fig. 1). All steps were performed for each dynamic flat panel detector (DFPD) image pair for each patient.

2.2.1. Step 1: Setting ROI and target probability map (TPM) positions on DRR images

One or more rectangular ROIs were set on 4D-DRR images derived from the 4DCT planning data sets. The ROIs were set at regions where the greatest displacement was noted in the images. Low contrast DFPD image prevented visualization of liver tumors; the ROIs in this study were placed to include sufficient tumor displacement region. DRR image pixel size and matrix size were 297 × 297 and 1.0 × 1.0 mm, respectively. The center of the ROI ($\boldsymbol{r} \in [u, v]$) was defined as:

$$\boldsymbol{r} = \arg\max_{(u,v)\in R} \sum_{t=0}^{T-2} \sum_{|\delta_u|\le W_r/2, |\delta_v|\le H_r/2} [I_{u+\delta_u, v+\delta_v, t+1} - I_{u+\delta_u, v+\delta_v, t}]^2 \quad (1)$$

where $I_{u,v,t}$ is the pixel value at location *(u, v)* on the *t* th frame of DRR image, *R* is the aggregate of image pixel positions, $W_r \times H_r$ is the size of the ROI, and *T* is a frame number from the 4DCT data set, generally 10. The center of the second and the following ROIs was set to avoid overlapping of ROIs. The total number of ROIs was $\boldsymbol{N_r}$ (=4 in this study). In this study, the values of W_r and H_r were 64 and 128 pixels, respectively, in order to include the tumor displacement region.

Next, a target probability map (TPM) position was defined at the center of the tumor position for respective respiratory phases on DRR images, with the TPM size the same as the ROI size (yellow[1] square in Fig. 1).

2.2.2. Step 2: Preparing training data

Training data were prepared from 4D-DRR images and the reference tumor positions. To increase the amount of training data, DRR images were calculated with randomly shifted CT positions (< ± 2 mm and ± 1 deg), whose values depend on the patient setup error, and tumor positions on the shifted CT positions were also projected onto the DRRs. The DRR images and the projected tumor positions were rotated (± 1 deg) and translated (data augmentation). DRR image pixel size was 1.0 × 1.0 mm. Pixel values of all DRR images were normalized within the range of 0–1. In this study, 1000 augmentation patterns were applied to 10 respiratory CT data sets, resulting in 10,000 training images.

The N_r subimages were clipped from each training image as an input vector for the network. A TPM, which is an output vector for the network, was also calculated by the following equation:

$$t_{u,v} = \exp\left[\frac{||\boldsymbol{z}_{u,v} - \boldsymbol{z}_{tumor}||^2}{2\sigma^2}\right] \quad (2)$$

where $t_{u,v}$ is a value of a TPM at location *(u, v)*, $\boldsymbol{z}_{u,v}$ is a vector at location *(u, v)* on a training image, $\boldsymbol{z}_{tumor}$ is a tumor position projected onto the image and $\sigma = \min(W_r, H_r)/8$.

2.2.3. Step 3: Network training: optimizing the parameters of the tumor tracking DNN

Tumor tracking DNN ($y(\boldsymbol{x}; \theta)$), where θ represents the parameters of the DNN, was trained to export a TPM by importing subimages within the ROIs, denoted by $\boldsymbol{x}$, by minimizing loss function (*L*).

$$L(\theta) = \frac{1}{2N}\sum_{i=1}^{N} ||y(\boldsymbol{x}^{(i)}; \theta) - \boldsymbol{t}^{(i)}||^2 + \frac{\lambda}{2}||\theta||^2 \quad (3)$$

where $\boldsymbol{x}^{(i)} \in \mathbb{R}^{W_r \times H_r \times N_r}$ is training data, which is composed of N_r subimages, clipped from the *i* th training image. $\boldsymbol{t}^{(i)} \in \mathbb{R}^{W_r \times H_r}$ is a TPM calculated by using the tumor position on the *i* th training image.

An optimization procedure was performed for 1000 epochs using adaptive moment estimation (ADAM) with a minibatch size of 16 images and weight decay (=λ in Eq. [3]) of 5.0×10^{-4} [16]. The encoder and decoder weights were all initialized as proposed by Glorot et al. [17]. A dropout layer with a rate of 0.1 was added to the last layers in the respective sets [18].

[1] For interpretation of color in Fig. 1, the reader is referred to the web version of this article.

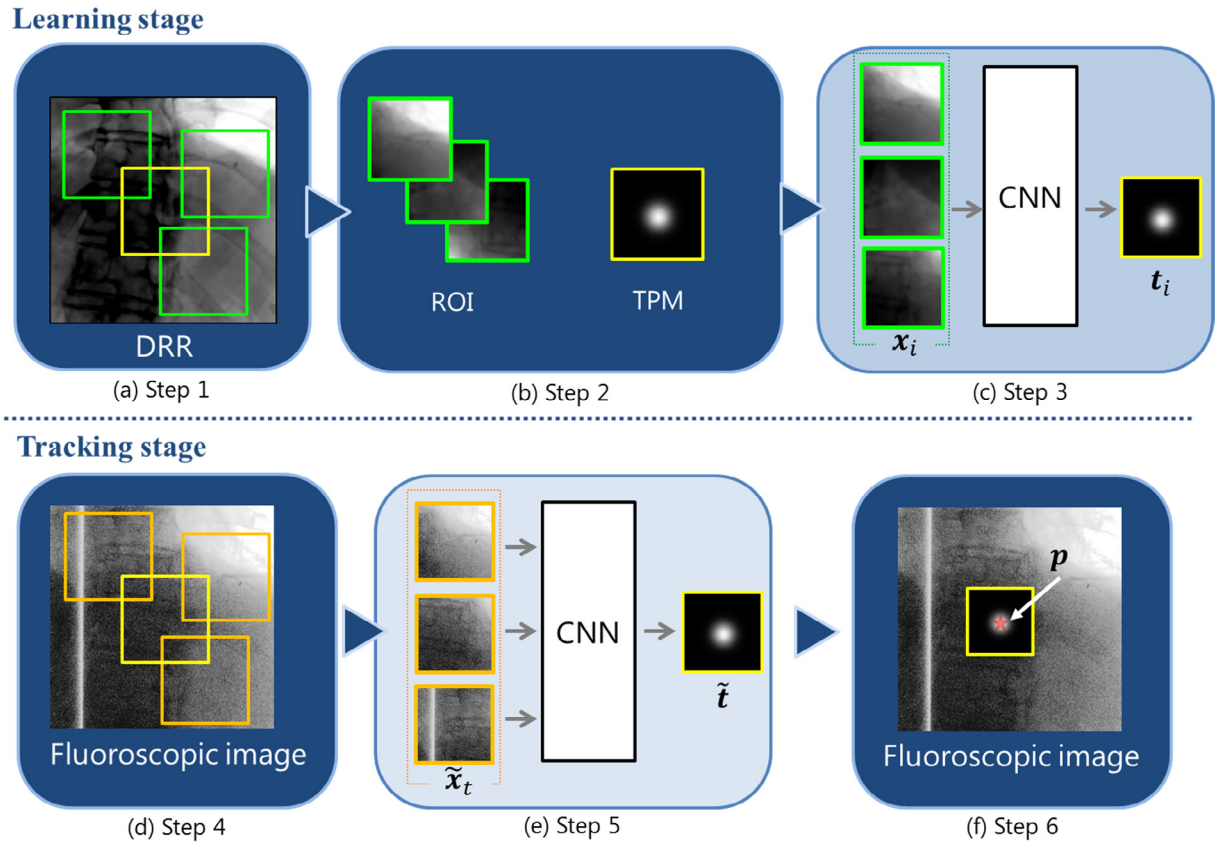


Fig. 1. Tracking algorithm procedures. (a) Step 1: ROIs were set on the DRR images. (b) Step 2: Training data were prepared from 4D-DRR subimages ($x^{(i)}$) and their TPMs based on the ground truth tumor positions ($t^{(i)}$). (c) Step 3: Parameters of the tumor tracking DNN were optimized (d) Step 4: Clipping of the subimages (x_t) of ROIs placed on DFPD images. (e) Step 5: The DNN generated TPM (t) (f) Step 6: A tumor position (p) on DFPD images was calculated from TPM. Abbreviations: DRR = digital reconstructed radiography. ROI = region of interest, TPM = target probability map, DNN = deep neural network, DFPD = dynamic flat panel detector.

2.2.4. Step 4: Clipping ROI images on dynamic flat panel detector (DFPD) images

DFPD image pixel size (0.388 mm), which was acquired during treatment, was resized to DRR image size (1.0 mm). Pixel values of the DFPD images were normalized similarly to those of the training DRR images. Subimages were clipped on ROIs (ROI_1 to ROI_{Nr}) from the same positions defined in Step 1 of the resized fluoroscopic images (Fig. 1d).

2.2.5. Step 5: Generating a TPM

Subimages from ROIs on the resized image were assigned to respective channels ($\boldsymbol{x} \in \mathbb{R}^{W_r \times H_r \times N_r}$), and a TPM ($\boldsymbol{t}$) was estimated by importing the subimages (ROIs) into the DNN ($\boldsymbol{t} = y(\boldsymbol{x}; \hat{\theta})$) (Fig. 1e), where $\hat{\theta}$ is the parameter that results in the best loss function approximation (Eq. (3)).

2.2.6. Step 6: Calculating a tumor position

A tumor position on DFPD image ($\boldsymbol{p} \in \mathbb{R}^2$) was derived by positional weighted averaging of a TPM ($\boldsymbol{t}$).

$$\boldsymbol{p} = \frac{1}{M} \sum_{(u,v) \in R_T} \boldsymbol{t}_{u,v} \begin{pmatrix} u \\ v \end{pmatrix} \tag{4}$$

where R_T is the set of all the pixel positions in the TPM and M is expressed as follows:

$$M = \sum_{(u,v) \in R_T} \boldsymbol{t}_{u,v} \tag{5}$$

In this study, tumor positions were calculated on two DFPD images each, which were obtained by a pair of fluoroscopic imaging systems. Then the tumor position in 3D space was derived from the 2D tumor positions.

Generally, respiratory-induced organ motion in the superior-inferior (SI) direction is larger than that in the lateral direction. Since ROI positions included large displacement image regions, as described in the previous section, image variation in the ROI would be mainly in the SI direction, making it difficult to resolve a tumor position from a displacement in the lateral direction.

To solve this problem, a tumor position subject to displacement in the lateral direction was corrected by the linear regression model in between tumor position in SI and lateral directions (lateral correction). A linear regression model from the SI to lateral direction of a tumor position on the planning 4DCT was designed:

$$y_t = ax_t + b,\ t = 1, \cdots, T \tag{6}$$

where C is the number of frames of 4DCT data sets, x_t and y_t are the SI and lateral directions of tumor positions where the tumor position on the 4DCT was projected onto a DRR, respectively, and the model parameters a and b are the solutions of these simultaneous linear equations. At the tracking stages, the lateral tumor position y is calculated by $y = ax + b$, where $\times$ is the SI tumor position estimated by Eq. (5).

2.3. Network architecture

2.3.1. Encoder and decoder network

Our tracking algorithm trained network parameters θ to generate a TPM using DRR images before treatment, and estimated tumor positions on DFPD images during treatment. Since patient position was registered to the reference position defined on the planning CT, DRR and DFPD images included the same anatomical information; however,

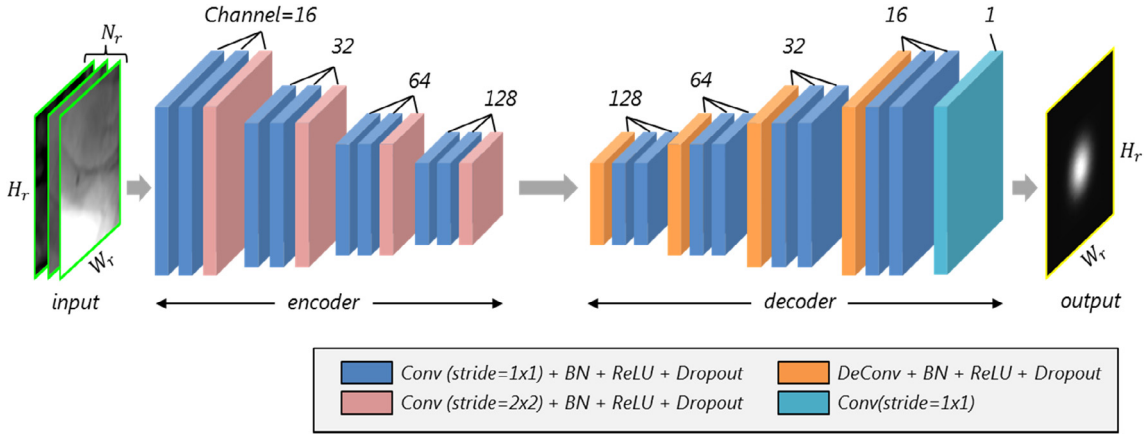


Fig. 2. Network structure for tumor tracking based on a convolutional autoencoder. *Nr*, *Hr* and *Wr* are the number of images, image height and image width, respectively. Abbreviations: Conv = convolution; BN = batch normalization; ReLU = rectified linear units; Deconv = deconvolution.

the quality of these images was different.

To solve this problem, our DNN was based on a convolutional auto encoder network (CAE). A similar approach was proposed as a Semantic Segmentation network (SegNet), which imported 3-channel color images that it processed and exported as semantic-labeled images [19]. The characteristics of our DNN led to its exporting not tumor position directly, but a TPM. The CAE reduced an input image into lower-dimensional data (Fig. 2) consisting of the 12 sets convolutional (Conv) + batch normalization (BN) [20] + rectified linear unit (ReLU) + Dropout layers. The stride size of each first and second convolutional layer was 1 × 1 pixel and that of the third one was 2 × 2 pixels. The kernel size of the convolutional layer was 3 × 3 pixels. The number of channels of the convolutional layer was 16, 32, 64, and 128 for the three respective sets.

The decoder network consisted of four sets of the following layers: deconvolutional (Deconv) + BN + ReLU layers, 2 × (Conv + BN + ReLU + Dropout) layers. The deconvolutional layer doubled the size of the receptive field and decreased the number of channels by half [21]. By doing this, the last ReLU layer outputs the same receptive field as that of the input data. Finally, a convolutional layer with kernel size of 1 × 1 pixel was added to output a single channel of data (TPM). By doing this, a feature map can be derived without any image quality dependency in between DRR and DFPD images. Next, the decoder network calculated a TPM using a feature map.

2.4. Evaluation

2.4.1. Tracking accuracy

Reference tumor positions on each DFPD image were calculated by regression-based 2D-2D rigid image registration [23] using the 4D-DRR with CTV counters and DFPD images, and were modified manually by both the certificated oncologist and medical physicist on the basis of the 4D-DRR. Tracking accuracy was evaluated as the Euclidian distance between the reference and calculated tumor positions. We compared the tracking accuracy between that with and without lateral correction based on a linear regression model.

2.4.2. Computation time

Computation time for the training (steps 1–3) and tracking (step 4–6) stages was averaged over all fluoroscopic images. Our tracking algorithm was programmed using the C++ program language (Microsoft Visual Studio 2012®, Microsoft, Redmond WA, USA). It works under a Windows 7 environment and is installed on a workstation (Intel Xeon® CPU@3.6 GHz, 32 GB physical memory, Hewlett-Packard, Inc., Palo Alto CA, USA) and a single GPU on a board (NVIDIA TitanX ©, NVIDIA Corporation, Santa Clara, CA, USA) with cuda8.0 and cuDNN v6 acceleration, which is equipped with 3072 compute unified device architecture (CUDA) cores, and has 12 GB of memory.

3. Results

3.1. Tracking accuracy

Tumor tracking positions in the SI direction fluctuated slightly as a function of time (liver 1); however, these were close to the reference (Fig. 3a). Tumor displacements in the left-right (LR) and anterior-posterior (AP) directions were < 1 mm. The tracking positional accuracy without the lateral correction caused 1.55 ± 0.19 mm and 0.25 ± 0.19 mm displacement in the LR and AP directions, respectively (Fig. 3b, c and Table 1). The lateral correction, however, improved the tracking accuracy to 0.18 ± 0.06 mm and 0.03 ± 0.01 mm in the LR and AP directions, respectively. The positional error expressed in Euclidean distance was 1.62 ± 0.20 mm and 0.36 ± 0.17 mm without and with the lateral correction, respectively (Fig. 3d, Table 2).

In another case (liver 2), where the worst case of tracking error was in the SI direction, the tumor tracking accuracy was degraded around the 120th frame (Fig. 4a). To clarify this, we visualized the TPM and overlaid it over the DFPD (Fig. 4b and c). The tumor tracking was close to the reference at T1, and the TPM showed a Gaussian morphology (Fig. 4b). However, the tracking positional error was approximately 5 mm at frame 2, because the TPM was losing the Gaussian morphology compared to that at frame 1 and a higher value was observed around the right side.

Tracking positional accuracy without the correction averaged over all patients was 2.18 ± 0.89 mm (95% CI: 2.30 mm) in Euclidian distance. By applying the lateral correction, those were improved to 1.64 ± 0.73 mm (95% CI: 1.74 mm). Accuracy for liver cases with the lateral correction (1.37 ± 0.81 mm) was better than that for lung cases (1.90 ± 0.65 mm). Results are summarized in Tables 1 and 2.

3.2. Computation time

Training time for DNN was approximately 11 h for a single DFPD direction. Computation time for tumor tracking with a pair of DFPDs averaged over all patients was 39.8 ± 3.7 ms per frame. Those of liver and lung cases were almost the same (39.7 ± 4.0 ms per frame for liver case, 39.9 ± 3.4 ms per frame for lung case) because of the use of the same ROI size for all patients. Results are summarized in Table 1.

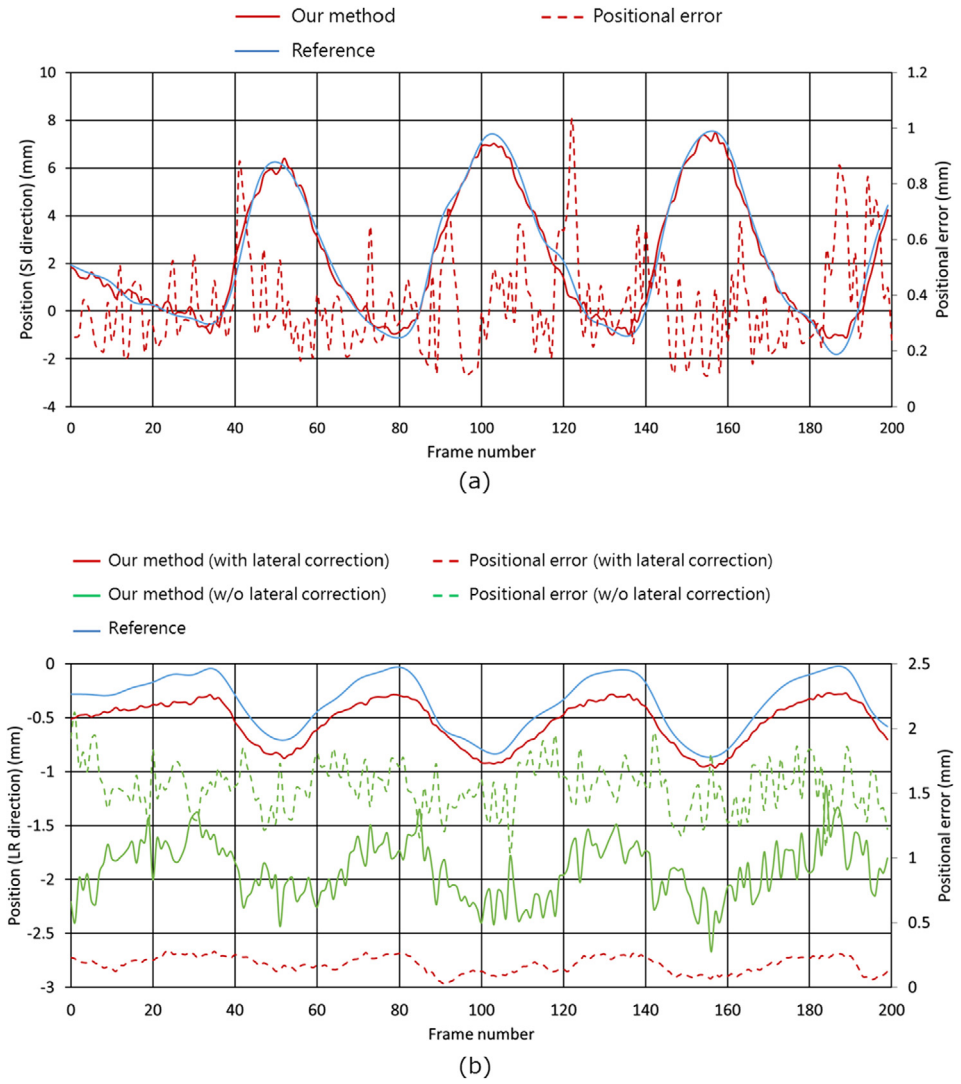


Fig. 3. Tumor tracking results using our method with and without the lateral correction and reference positions for a liver patient (liver 1): (a) superior-inferior (SI), (b) left-right (LR), and (c) anterior-posterior (AP) directions. (d) Tracking positional error in 3D space (tracking position minus reference position).

4. Discussion

We evaluated the accuracy of our DNN-based markerless tracking system using five liver and five lung cases. Our algorithm achieved positional accuracy of 1.64 ± 0.73 mm in Euclidian distance. Although the positional error seems large, it corresponds to 0.94 mm in each of the 3 directions. Moreover, because the computation time was < 40 ms for a pair of fluoroscopic images, the algorithm can be performed at a 15 fps imaging rate.

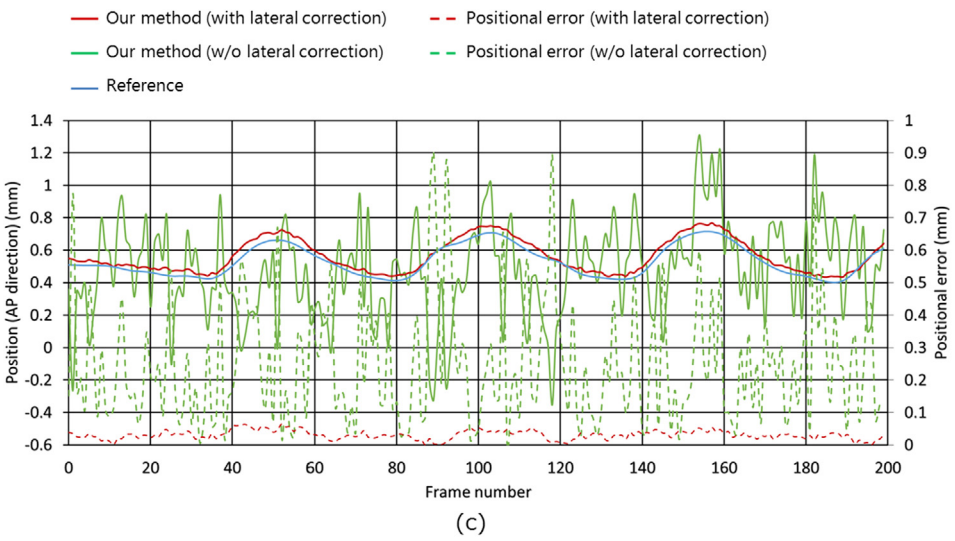
Reports of other markerless tracking systems required tumor positions on respective fluoroscopic images [7,8]. The manual input of tumor position before irradiation took a large amount of time, which would degrade treatment throughput. Our algorithm does not require the input of target positions on fluoroscopic images, but instead uses planning 4DCT and target contour input from the treatment planning system, obviating the need for the extra procedure.

Shieh et al. reported tracking accuracy of 1.6–2.9 mm in 3D Euclidian distance, which was inferior to our lung tracking accuracy (0.99–2.6 mm, Table 1). Moreover, the application of their algorithm to liver cases is questionable, because the lower contrast might obscure liver tumor features.

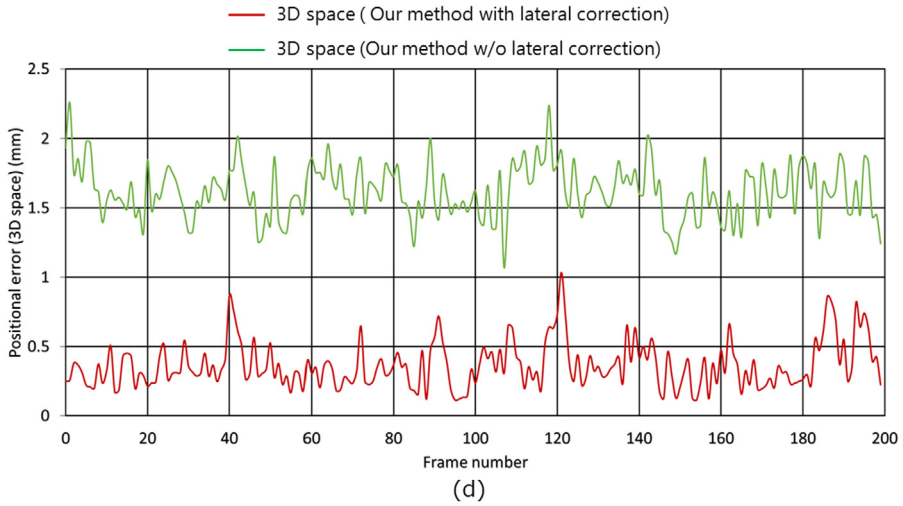
As described in Materials and Methods, since tumor displacement in the lateral direction was so small, tracking accuracy in the lateral direction was markedly degraded if the lateral correction was not applied. To solve this, we feel it is reasonable to use a linear regression model since there is strong correlation information of the tumor position in SI, LR and AP directions [24]. As a similar approach, Cervino et al. (2010) developed indirect markerless lung tumor tracking using diaphragmatic motion [12]. Tracking accuracy was 1.2 mm in the SI direction for lung patients, or in other words lower than our results (0.77 mm in Table 1 for all lung patients). However, our lateral correction used tumor position in the SI direction, and not diaphragm position. The lateral correction can therefore be applied to all patients, if the diaphragm is not visualized on DFPD images.

We expected that lung tracking accuracy would be higher than liver because the lung tumors were easier to see in 3D space (Table 2). Surprisingly, however, liver tracking accuracy was higher than that of lung. This is because the lateral correction used tumor displacement in the SI direction, and liver tumor displacement was generally in that direction, whereas lung tumor displacement was in three dimensions (different trajectory in exhale and inhale) [25]. The lateral correction was considered to represent SI motion only, and therefore improved liver tracking accuracy more than that of lung.

As an example of a low tracking accuracy case (liver 2) in the SI direction (Table 1), the TPM morphology was degraded from a Gaussian shape because the input DFPD image was so different from the



(c)



(d)

Fig. 3. (*continued*)

Table 1
Tracking errors in the SI, LR and AP directions.

	Tracking error [mm]									
	SI direction		LR direction				AP direction			
			w/o LATC		with LATC		w/o LATC		with LATC	
Patient no	mean ± SD	95% CI	mean ± SD	95% CI	mean ± SD	95% CI	mean ± SD	95% CI	mean ± SD	95% CI
Liver1	0.29 ± 0.20	0.32	1.55 ± 0.19	1.58	0.18 ± 0.06	0.19	0.25 ± 0.19	0.27	0.03 ± 0.01	0.03
Liver2	1.60 ± 1.83	1.86	0.55 ± 0.39	0.60	0.41 ± 0.34	0.46	0.53 ± 0.44	0.59	0.30 ± 0.31	0.35
Liver3	1.10 ± 0.61	1.19	0.32 ± 0.27	0.36	1.52 ± 0.54	1.59	1.46 ± 0.90	1.58	0.43 ± 0.25	0.46
Liver4	1.20 ± 0.62	1.29	0.76 ± 0.28	0.79	0.20 ± 0.08	0.21	1.36 ± 0.46	1.42	0.29 ± 0.12	0.31
Liver5	1.35 ± 1.05	1.50	1.70 ± 0.29	1.74	0.40 ± 0.21	0.43	1.52 ± 0.79	1.63	0.41 ± 0.18	0.44
Lung1	0.57 ± 0.14	0.59	0.20 ± 0.09	0.21	1.24 ± 0.09	1.25	0.75 ± 0.18	0.78	1.00 ± 0.14	1.02
Lung2	0.94 ± 0.95	1.07	1.30 ± 0.76	1.41	1.70 ± 0.76	1.8	2.31 ± 1.77	2.55	0.69 ± 0.57	0.77
Lung3	0.91 ± 0.60	0.99	1.07 ± 0.72	1.17	0.97 ± 0.65	1.06	2.57 ± 1.76	2.82	2.11 ± 1.08	2.26
Lung4	0.91 ± 0.62	1.00	0.88 ± 0.71	0.97	0.22 ± 0.18	0.24	0.54 ± 0.31	0.58	0.26 ± 0.22	0.29
Lung5	0.46 ± 0.39	0.52	0.75 ± 0.37	0.81	0.88 ± 0.34	0.93	2.28 ± 0.85	2.39	1.55 ± 0.45	1.62
All livers	1.11 ± 0.86	1.23	0.97 ± 0.28	1.01	0.54 ± 0.25	0.57	1.02 ± 0.56	1.10	0.29 ± 0.17	0.32
All lungs	0.76 ± 0.54	0.83	0.84 ± 0.54	0.91	1.00 ± 0.40	1.06	1.69 ± 0.97	1.82	1.12 ± 0.49	1.19
All patient	0.93 ± 0.70	1.03	0.91 ± 0.41	0.96	0.77 ± 0.32	0.82	1.36 ± 0.77	1.46	0.71 ± 0.33	0.75

Abbreviations: SI = superior-inferior, LR = left-right, AP = anterior-posterior, SD = standard deviation, CI = confidence interval.

Table 2
Tracking errors in 3D space and computation time per frame.

	Tracking error [mm]				Computation time [ms]
	3D space				
	w/o LATC		with LATC		
Patient no	mean ± SD	95% CI	mean ± SD	95% CI	mean ± SD
Liver1	1.62 ± 0.20	1.65	0.36 ± 0.17	0.39	38.0 ± 4.8
Liver2	1.96 ± 1.74	2.20	1.73 ± 1.84	1.98	40.2 ± 5.2
Liver3	2.00 ± 0.84	2.12	2.00 ± 0.58	2.08	39.8 ± 3.5
Liver4	2.08 ± 0.42	2.14	1.27 ± 0.60	1.35	40.5 ± 2.2
Liver5	2.93 ± 0.50	3.00	1.48 ± 0.86	1.60	39.8 ± 3.0
Lung1	0.98 ± 0.14	1.00	1.70 ± 0.11	1.71	40.4 ± 2.4
Lung2	3.14 ± 1.66	3.37	2.32 ± 0.81	2.43	40.5 ± 3.1
Lung3	3.10 ± 1.72	3.33	2.60 ± 1.18	2.77	40.2 ± 3.3
Lung4	1.52 ± 0.75	1.63	0.99 ± 0.65	1.08	39.3 ± 3.7
Lung5	2.49 ± 0.89	2.61	1.90 ± 0.48	1.97	39.0 ± 4.0
All livers	2.12 ± 0.74	2.22	1.37 ± 0.81	1.48	39.7 ± 4.0
All lungs	2.25 ± 1.03	2.39	1.90 ± 0.65	1.99	39.9 ± 3.4
All patient	2.18 ± 0.89	2.30	1.64 ± 0.73	1.74	39.8 ± 3.7

Abbreviations: LATC = lateral correction, SD = standard deviation, CI = confidence interval.

training image data due to deep breathing. To prevent this, we increased the number of image variations by data augmentation. However, if the patient's position at treatment was greatly altered from that on the reference 4DCT (tumor and bony structures), the tracking accuracy could be degraded. To solve this, it will be useful to integrate the interlock function by checking similarity between the TPM shape and that of the training data.

Our DNN achieved a real-time markerless tumor tracking (< 40 ms), and is available for use with 15 fps imaging (67 ms). Since DFPD image size was reduced to that of the DRR images, tracking accuracy could be decreased but computation time was shortened. Use of large image sizes requires high computation GPUs to increase tracking accuracy in real time.

Our DNN also achieved shortening of the treatment time for gating. Mori et al. (2018) reported that nearly 30% of the total patient setup time was longer with respiratory gating than without gating when using markerless tumor tracking based on multi-template matching [26]. The templates were made by applying 2-dimensional image registration between 4D-DRR and DFPD and checked by oncologists. By contrast, our method does not need templates, and therefore has a shortened setup time.

While the network training time was approximately 1 day (11 h × 2 directions), use of two computers or a multi-GPU system would shorten this to within half a day. If network training were done using a workstation which was not used for treatment, our DNN would not degrade treatment throughput.

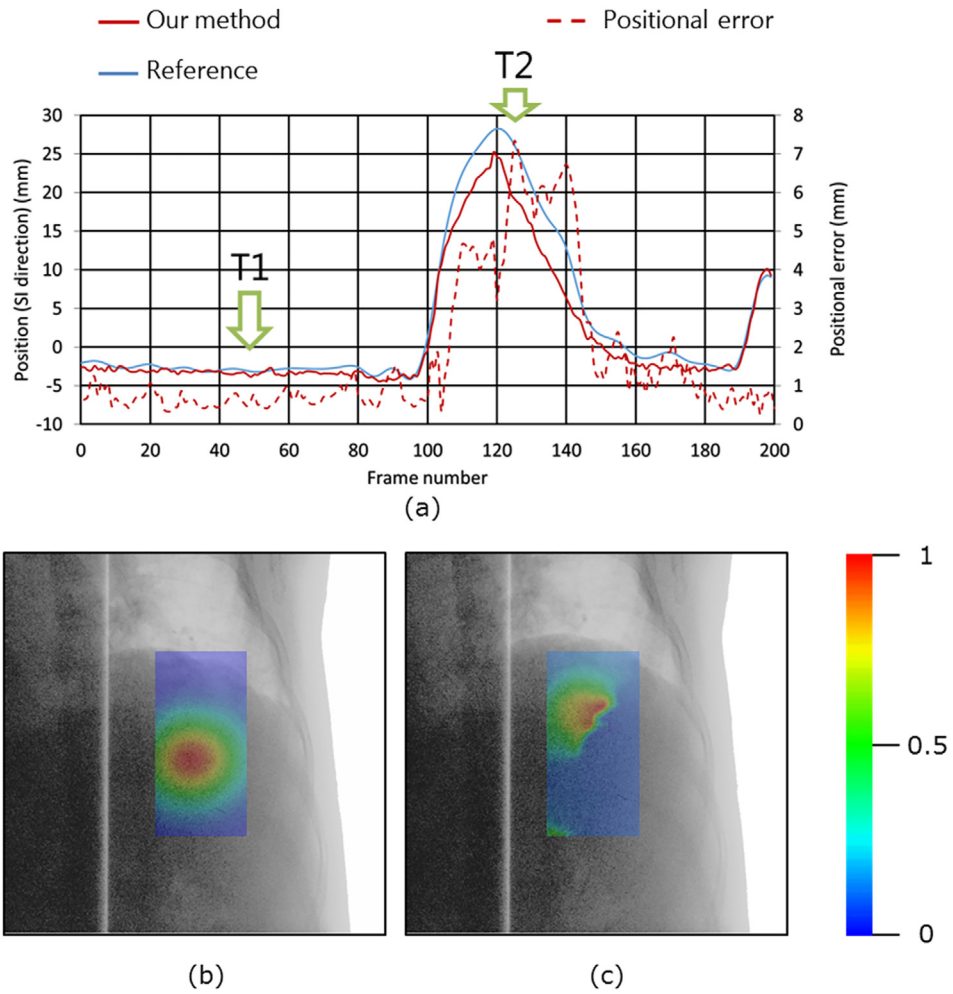


Fig. 4. An example of a difficult liver case (liver 2). (a): Tumor positions along the superior-inferior direction for the reference and our method (left axis). Positional error was also plotted (our method minus the reference) (right axis). TPM overlaid on DFPD image at (b) time = T1 (50th frame) and (c) time = T2 (130 th frame) for DFPD1 image. Abbreviations: TPM = tumor positional map, DFPD = dynamic flat panel detector.

A few limitations of this study warrant mention. First, our DNN was trained using the treatment planning 4DCT data, which were obtained before irradiation. The DNN might degrade tracking accuracy when patient organ position and size change between fractions, even though the DNN was trained with the transformed image data (data augmentation). Moreover, to increase tracking accuracy, our DNN considered the relationship between tumor and surrounding organ positions by importing subimages around the tumor region. However, this might degrade tracking accuracy when tumor and organ positions are subject to interfractional changes. Carbon-ion beam treatment requires fewer treatment fractions than proton and photon beams. However, as it took generally a few days to a week between planning CT acquisition and treatment, interfractional change was not negligible.

Second, the machine learning approach including DNN calculated a model file using training data and optimized large parameters automatically. The integration of DNN-based markerless tracking into the clinical application is useful because it helps smooth throughput; however, in cases in which this DNN-based markerless tracking did not work well, the treatment could not continue because there was no space to modify the model file to estimate accurate tumor positions. We are developing new techniques to solve this problem.

Third, because it is hard to input the exact reference tumor position on DFPD images for patient studies, we used regression-based 2D-2D rigid image registration. Even with the use of 4D-DRR images which include organ deformation, rigid image registration might still have simplified the deformation. Therefore, when required, a certificated oncologist and medical physicist modified the reference position by accounting for organ deformation. This limitation remains in any studies using patient data.

5. Conclusion

We describe DNN-based markerless tumor tracking using 4DCT data sets. A major advantage of our DNN system is its improved tracking accuracy, especially in the lateral direction via the assessment of position changes of both the tumor and surrounding tissues and organs. Allowing that we evaluated only a limited number of liver and lung patients, we believe our results will be useful in the progress of markerless tumor tracking technology.

Acknowledgments

Mr. Hirai, Dr. Sakata and Mr. Tanizawa are employed by the Toshiba Corporation, Kawasaki, Japan. We thank Libby Cone, MD, MA, from DMC Corp. (www.dmed.co.jp < http://www.dmed.co.jp/ >) for editing drafts of this manuscript.

References

[1] Jakel O, Karger CP, Debus J. The future of heavy ion radiotherapy. Med Phys 2008;35:5653–63.
[2] Riboldi M, Orecchia R, Baroni G. Real-time tumour tracking in particle therapy: technological developments and future perspectives. Lancet Oncol 2012;13:e383–91.
[3] Kubo HD, Hill BC. Respiration gated radiotherapy treatment: a technical study. Phys Med Biol 1996;41:83–91.
[4] Meschini G, Seregni M, Pella A, Ciocca M, Fossati P, Valvo F, et al. Evaluation of residual abdominal tumour motion in carbon ion gated treatments through respiratory motion modelling. Phys Med Biol 2017;34:28–37.
[5] Shirato H, Harada T, Harabayashi T, Hida K, Endo H, Kitamura K, et al. Feasibility of insertion/implantation of 2.0-mm-diameter gold internal fiducial markers for precise setup and real-time tumor tracking in radiotherapy. Int J Radiat Oncol Biol Phys 2003;56:240–7.
[6] Giraud P, Garcia R. Respiratory gating for radiotherapy: main technical aspects and clinical benefits. Bull Cancer 2010;97:847–56.
[7] Mori S, Karube M, Shirai T, Tajiri M, Takekoshi T, Miki K, et al. Carbon-ion pencil beam scanning treatment with gated markerless tumor tracking: an analysis of positional accuracy. Int J Radiat Oncol Biol Phys. 2016;95:258–66.
[8] Cui Y, Dy JG, Sharp GC, Alexander B, Jiang SB. Multiple template-based fluoroscopic tracking of lung tumor mass without implanted fiducial markers. Phys Med Biol 2007;52:6229–42.
[9] Li R, Lewis JH, Cervino LI, Jiang SB. A feasibility study of markerless fluoroscopic gating for lung cancer radiotherapy using 4DCT templates. Phys Med Biol 2009;54:N489–500.
[10] Zhang X, Homma N, Ichiji K, Abe M, Sugita N, Takai Y, et al. A kernel-based method for markerless tumor tracking in kV fluoroscopic images. Phys Med Biol 2014;59:4897–911.
[11] Zhang X, Homma N, Ichiji K, Takai Y, Yoshizawa M. Tracking tumor boundary in MV-EPID images without implanted markers: a feasibility study. Med Phys 2015;42:2510–23.
[12] Cervino LI, Jiang Y, Sandhu A, Jiang SB. Tumor motion prediction with the diaphragm as a surrogate: a feasibility study. Phys Med Biol 2010;55:N221–9.
[13] Cervino LI, Chao AK, Sandhu A, Jiang SB. The diaphragm as an anatomic surrogate for lung tumor motion. Phys Med Biol 2009;54:3529–41.
[14] Sharp GC, Li R, Wolfgang J, Chen G, Peroni M, Spadea MF, et al. Plastimatch—an open source software suite for radiotherapy image processing. ICCR 2010: XVIth international conference on the use of computers in radiation therapy. Amsterdam, Netherlands. 2010.
[15] Sharp GC, Kandasamy N, Singh H, Folkert M. GPU-based streaming architectures for fast cone-beam CT image reconstruction and demons deformable registration. Phys Med Biol 2007;52:5771–83.
[16] Kingma D, Ba J. Adam: a method for stochastic optimization. Proceedings of the 3rd international conference on learning representations (ICLR). 2015.
[17] Glorot X, Bengio Y. Understanding the difficulty of training deep feedforward neural networks. Society for Artificial Intelligence and Statistics; 2010.
[18] Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov N. Dropout: a simple way to prevent neural networks from overfitting. J Mach Learn Res 2014;15:1929–58.
[19] Badrinarayanan V, Kendall A, Cipolla R. SegNet: a deep convolutional encoder-decoder architecture for image segmentation. IEEE Trans Pattern Anal Mach Intell 2017;39:2481–95.
[20] Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift. Proceedings of the 32nd international conference on machine learning, PMLR. 2015. p. 448–56.
[21] Noh H, Hong S, Han B. Learning deconvolution network for semantic segmentation. IEEE international conference on computer vision (ICCV). 2015. p. 1520–8.
[22] Shieh CC, Caillet V, Dunbar M, Keall PJ, Booth JT, Hardcastle N, et al. A Bayesian approach for three-dimensional markerless tumor tracking using kV imaging during lung radiotherapy. Phys Med Biol 2017;62:3065–80.
[23] Sakata Y, Hirai R, Taguchi Y, Mori S. Marker-less tumor tracking for lung cancer by tumor image pattern learning. Int J Radiat Oncol 2016;96. E651-E651.
[24] Seppenwoolde Y, Shirato H, Kitamura K, Shimizu S, Van Herk M, Lebesque JV, et al. Precise and real-time measurement of 3D tumor motion in lung due to breathing and heart-beat, measured during radiotherapy. Int J Radiat Oncol Biol Phys 2002;53:822–34.
[25] Mori S, Endo M, Komatsu S, Yashiro T, Kandatsu S, Baba M. Four-dimensional measurement of lung tumor displacement using 256-multi-slice CT-scanner. Lung Cancer (Amsterdam, Netherlands) 2007;56:59–67.
[26] Mori S, Takei Y, Shirai T, Hara Y, Furukawa T, Inaniwa T, et al. Scanned carbon-ion beam therapy throughput over the first 7 years at National Institute of Radiological Sciences. Phys Med 2018;52:18–26.