IOP Publishing Phys. Med. Biol. 65 (2020) 085014 https://doi.org/10.1088/1361-6560/ab79c5

Physics in Medicine & Biology

IPEM Institute of Physics and Engineering in Medicine

RECEIVED
29 October 2019

REVISED
12 February 2020

ACCEPTED FOR PUBLICATION
25 February 2020

PUBLISHED
23 April 2020

PAPER

A machine learning-based real-time tumor tracking system for fluoroscopic gating of lung radiotherapy

Yukinobu Sakata[1], **Ryusuke Hirai**[1], **Kyoka Kobuna**[1], **Akiyuki Tanizawa**[1] **and Shinichiro Mori**[2]

[1] Corporate Research and Development Center, Toshiba Corporation, Kanagawa, Japan
[2] Research Center for Charged Particle Therapy, National Institute of Radiological Sciences, Chiba, Japan

E-mail: yuki.sakata@toshiba.co.jp

Keywords: markerless tumor tracking, machine learning, radiotherapy, real-time imaging

Abstract

To improve respiratory-gated radiotherapy accuracy, we developed a machine learning approach for markerless tumor tracking and evaluated it using lung cancer patient data. Digitally reconstructed radiography (DRR) datasets were generated using planning 4DCT data. Tumor positions were selected on respective DRR images to place the GTV center of gravity in the center of each DRR. DRR subimages around the tumor regions were cropped so that the subimage size was defined by tumor size. Training data were then classified into two groups: positive (including tumor) and negative (not including tumor) samples. Machine learning parameters were optimized by the extremely randomized tree method. For the tracking stage, a machine learning algorithm was generated to provide a tumor likelihood map using fluoroscopic images. Prior probability tumor positions were also calculated using the previous two frames. Tumor position was then estimated by calculating maximum probability on the tumor likelihood map and prior probability tumor positions.

We acquired treatment planning 4DCT images in eight patients. Digital fluoroscopic imaging systems on either side of the vertical irradiation port allowed fluoroscopic image acquisition during treatment delivery. Each fluoroscopic dataset was acquired at 15 frames per second. We evaluated the tracking accuracy and computation times.

Tracking positional accuracy averaged over all patients was 1.03 ± 0.34 mm (mean $\pm$ standard deviation, Euclidean distance) and 1.76 ± 0.71 mm (95^{th} percentile). Computation time was 28.66 ± 1.89 ms/frame averaged over all frames. Our markerless algorithm successfully estimated tumor position in real time.

1. Introduction

Radiation therapy is one of the main treatments for cancer. To deliver radiation to a target tumor accurately, very accurate and repeatable patient positioning is necessary (Jakel, Karger and Debus 2009, Mori, Shibayama, Tanimoto, Kumagai, Matsuzaki, Furukawa, Inaniwa, Shirai, Noda, Tsuji and Kamada 2012) because tumors, particularly in the thoracoabdominal region, move with patient respiration; hence the beam must be accurately controlled to follow changes in tumor position over time. There are two well-known methods for estimating tumor position: external surrogate motion tracking and internal tumor tracking. Kubo and Hill (Kubo and Hill 1996) integrated external gating into proton beam therapy using a motion sensor on the surface of the abdomen. However, it has been suggested that the correlation between the observed external respiratory signal and actual internal tumor motion was insufficiently accurate (Meschini, Seregni, Pella, Ciocca, Fossati, Valvo, Riboldi and Baroni 2017).

Shirato *et al* (Shirato, Harada, Harabayashi, Hida, Endo, Kitamura, Onimaru, Yamazaki, Kurauchi, Shimizu, Shinohara, Matsushita, Dosaka-Akita and Miyasaka 2003) developed fiducial marker tracking with fluorscopy. This technique solved the gating problem by directly detecting implanted markers. However, in order to estimate the tumor position from the marker position, the marker must be implanted near the

tumor but this may, depending on tumor position, carry the risk of pneumothorax (Li, Lewis, Cervino and Jiang 2009). To solve these problems, several other methods have proposed tracking tumors without fiducial markers (Cui, Dy, Sharp, Alexander and Jiang 2007a, 2007b, Rottmann, Aristophanous, Chen, Court and Berbeco 2010, Teo, Crow, Van Nest, Sasaki and Pistorius 2013, Bryant, Rottmann, Lewis, Mishra, Keall and Berbeco 2014, Zhang, Homma, Ichiji, Abe, Sugita, Takai, Narita and Yoshizawa 2014, Shieh, Keall, Kuncic, Huang and Feain 2015, Zhang, Homma, Ichiji, Takai and Yoshizawa 2015, Mori, Karube, Shirai, Tajiri, Takekoshi, Miki, Shiraishi, Tanimoto, Shibayama, Yasuda, Yamamoto, Yamada, Tsuji, Noda and Kamada 2016, Shieh, Caillet, Dunbar, Keall, Booth, Hardcastle, Haddad, Eade and Feain 2017).

One approach to markerless tumor tracking uses fluoroscopic image patterns. One method (Cui *et al* 2007a, 2007b) employed multi-template matching based on markerless tumor tracking using fluoroscopic images and is now in clinical use (Mori *et al* 2016). The templates are created from training fluoroscopic images taken during the patient setup session. However, it is necessary to manually input the tumor position on the template before treatment, which is very time consuming. Other methods to improve throughput using image pattern features employ either the mean-shift method or level set method (Zhang *et al* 2014, 2015). In addition, a method of tracking landmarks that are set based on the intensity variance by Normalized Cross-Correlation (Rottmann *et al* 2010) and another method of tracking tumors by optical flow have been proposed (Teo *et al* 2013). These methods require input of the tumor boundaries or the tumor position on the first frame, but this is difficult in the presence of motion artifact. To solve this, (Bryant *et al* 2014, Shieh *et al* 2015, Shieh *et al* 2017) proposed creating template images from treatment planning 4DCT images that include the position and contour of the tumor. However, the performance of this method was suboptimal.

To solve these problems, we developed a machine learning approach for markerless tumor tracking using 4DCT images. Before treatment, classifiers from training datasets are constructed to estimate the likelihood of the presence of a tumor given a certain image pattern. During treatment, features from the fluoroscopic images are sampled and classified to generate a tumor likelihood map, which shows the relationship between tumor position and tumor likelihood on the image. Tumor position is then estimated using the tumor likelihood map and prior information obtained from the tumor's motion. Here, we retrospectively evaluated the tracking accuracy of this proposed markerless method in eight lung cancer patients.

2. Methods and materials

2.1. Image acquisition

Treatment planning 4D-CT images were acquired under free-breathing conditions using a 320-slice CT (Aquilion One Vision©, Canon Medical Systems, Otawara, Japan). 4D-CT images were subdivided into 10 phases between T00 (peak inhalation) and T50 (around peak exhalation). CT section thickness was 1.0 mm. A board-certified oncologist manually contoured the CTV on the CT data at T50, and the contours at other respiratory phases were then automatically generated using B-Spline-based deformable image registration (Sharp, Kandasamy, Singh and Folkert 2007). The errors in the CTV contours created by deformable image registration were & 2 mm in Euclidean distance and the contours were checked by an oncologist and medical physicist (Mori, Karube, Yasuda, Yamamoto, Tsuji and Kamada 2017). Reference tumor positions in the respective phases were defined as the center of mass of the CTV.

Fluoroscopic images were acquired during treatment delivery using a dynamic flat panel detector (DFPD) (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 (figure 1) (Mori, Shirai, Takei, Furukawa, Inaniwa, Matsuzaki, Kumagai, Murakami and Noda 2012). Image size at the room isocenter is 298 × 298 mm, matrix size is 768 × 768 and pixel spacing is 0.388 mm × 0.388 mm. Fluoroscopy data were acquired at 15 frames per second (fps).

2.2. Tracking algorithm

The tumor tracking algorithm consists of three steps (figure 2); step 1 (learning classifier) is the training stage, which is performed before treatment, and steps 2 (likelihood map estimation) and 3 (tumor tracking) are the tracking stages. Tumor position in 3D space is expressed as $\mathbf{X} = (x, y, z)^t$, and on 2D flat panel images as $\mathbf{u} = (u, v)^t$, respectively, where the superscript t means transpose.

2.2.1. Step 1: Learning classifier

The classifier was trained using the 4DCT planning data of the target patient. Digitally reconstructed radiography (DRR) images of the chest for training data were generated from 4DCT planning data (figure 2(a)). Then, DRR subimages, consisting of small square regions of interest (ROIs), one baseline subimage encompassing the tumor, and others including regions containing partial to little or no tumor,

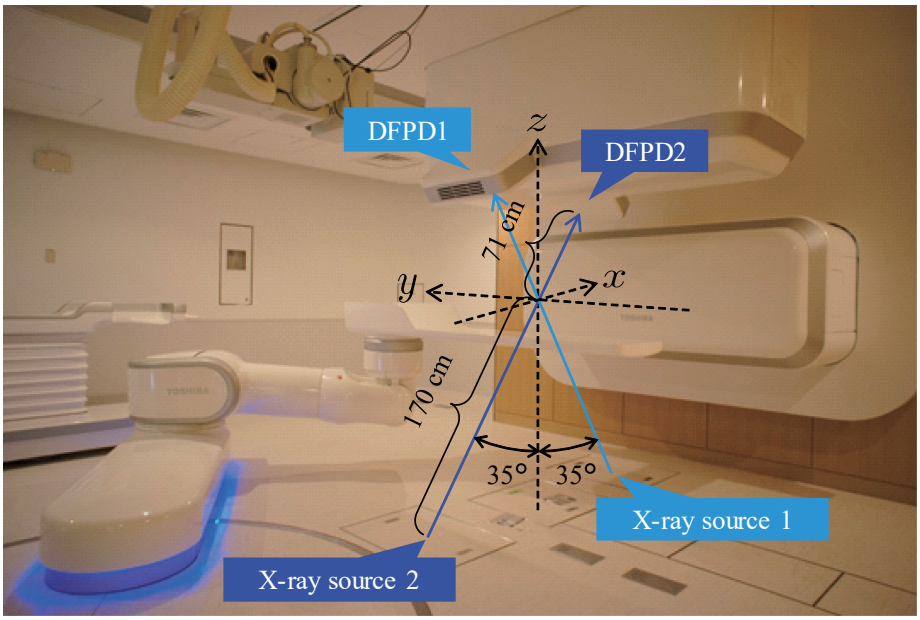

Figure 1. Geometry of the fluoroscopic imaging system in the treatment room. The room isocenter and source-image receptor distance are 170 cm and 241 cm, respectively. Image matrix size was 768 × 768 and physical pixel size was 0.388 × 0.388 mm.

were delineated. The subimage size was defined by clinical target volume (CTV) size as defined by treatment planning data (figure 2(b)). The cropped subimages were defined by the position coordinates *du*, *dv*, which represent the subimage position on DRR. The position of the baseline subimage that contains the tumor is (0.0, 0.0). Figure 3 demonstrates the scale and rotation angles of the subimages, with a baseline scale of 1.0 and a baseline rotation of 0°.

The position and size of the baseline subimages in other planes were automatically calculated from the tumor position and CTV, which were projected onto the DRR. By changing these parameters, multiple subimages could be cropped from one DRR. Subimage position, scale and rotation angle were changed by cropping (± 10 mm, 0.7–1.3, ± 10°) (data augmentation). Training data were then classified into two groups, namely positive and negative samples (figure 2(c)). The positive samples represent the center of the tumor in 2D. However, if a positive sample was derived from the center of the tumor only, the number of positive samples would be only one per DRR image, and therefore significantly smaller than the number of negative samples. This may not lend itself well to machine leaning model optimization due to the large difference in the number of positives and negatives. Positive samples were therefore derived from images sampled with a narrow margin around the tumor position (position < 3 mm, scale 0.9–1.1, rotation angle < 5°), and negative samples were made with subimages peripheral to the tumor (figure 4). The number of subimages was 1000 for each of 10 DRR phases for two sets of subimages, one positive and one negative, for a total of 1000 × 2 × 10 = 20000 subimages. The cropped subimages were converted into a feature vector by the method detailed in equations (1) and (2) below, and the classifier was learned.

Feature information on the subimages was assessed using intensity gradients to minimize the image quality differences between the digitally reconstructed radiographic (DRR) and fluoroscopic images. The intensity of an image at the image coordinate **u** was $I(\mathbf{u})$. The intensity gradient $G(\mathbf{u})$ was calculated as follows:

$$G(\mathbf{u}) = \begin{cases} g(\mathbf{u}) & (0 \le I_u) \\ g(\mathbf{u}) + \pi & (I_u < 0, 0 \le I_v) \\ g(\mathbf{u}) - \pi & (I_u < 0, I_v < 0) \end{cases} \tag{1}$$

$$g(\mathbf{u}) = \tan^{-1}\frac{I_v}{I_u},\ I_u = \frac{\partial I(\mathbf{u})}{\partial u},\ I_v = \frac{\partial I(\mathbf{u})}{\partial v} \tag{2}$$

Even in images with different intensity ranges, the intensity gradient direction will be close if the objects have the same shape (figure 5(a)). The correlation between the DRR baseline subimage cropped to contain the tumor position and fluoroscopic images containing the tumor were evaluated based on the intensity gradient direction (inner product between unit vectors) and the intensity (zero means Normalized Cross Correlation) (figure 5(b)). In the intensity gradient direction, the correlation was maximum at the tumor position,

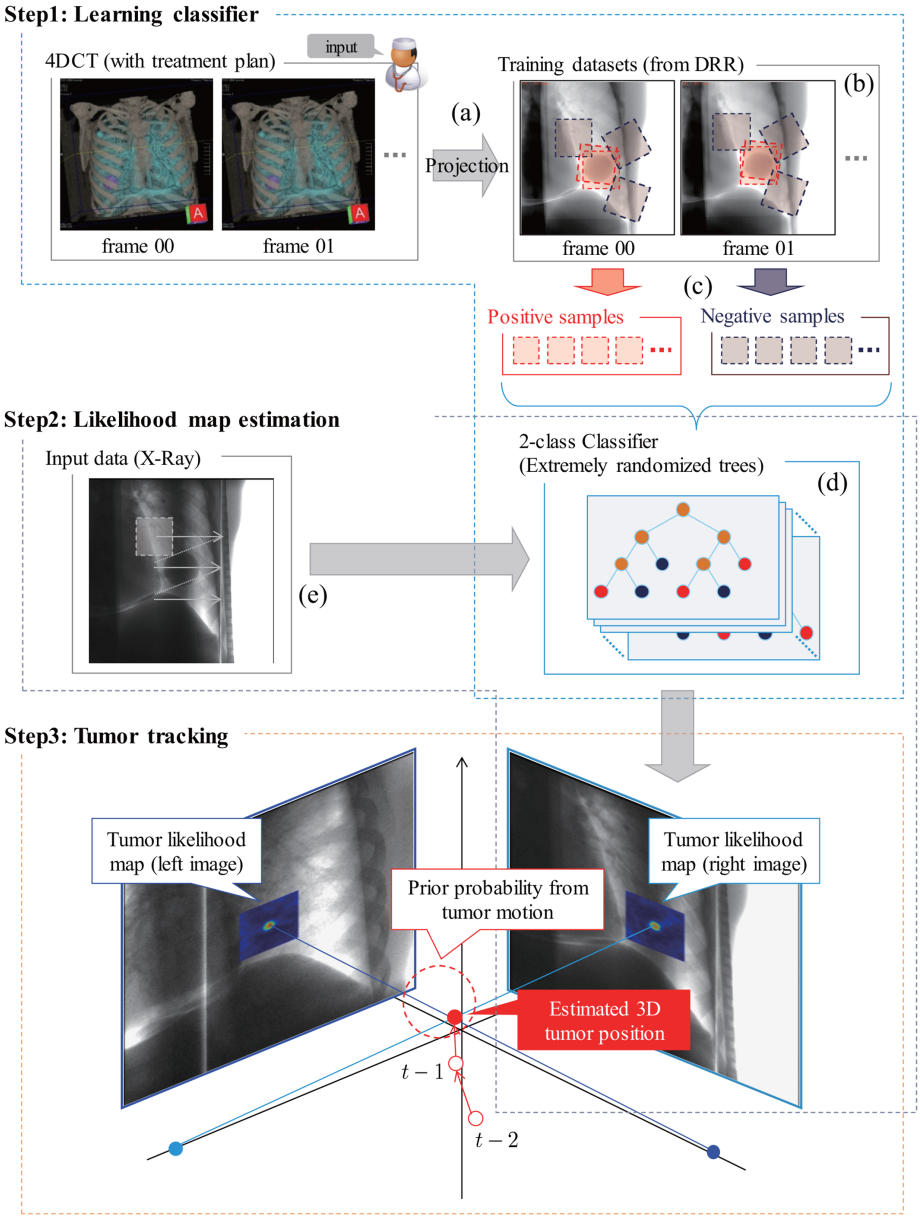

Figure 2. Step 1 (learning classifier): (a) DRR images as a function of respiratory phase (4D-DRR) were calculated using 4DCT data. (b) Two types of 4D-DRR subimages around tumor regions were cropped. (c) These images were classified into two groups; positive (including tumor) and negative (not including tumor). (d) Classifier parameters were then optimized by the extremely randomized tree method. Step 2 (likelihood map estimation): Tumor likelihood maps were generated with a classification approach using a sliding subimage (marked in dotted white square in figure 2(e)) to include the entire search region. Step 3 (tumor tracking): Tumor position was estimated by the tumor likelihood map and prior probability obtained from the tumor motion. Tumor position in 3D space can be derived using paired x-ray fluoroscopic images.

whereas the intensity was maximum at a position slightly deviated from the tumor position (figures 5(c) and (d)). In this way, by using the intensity gradient direction to reduce the image quality difference between the DRR and fluoroscopic images, the tumor position on the fluoroscopic image can be detected using the image pattern learned from the DRR.

Machine learning parameters were optimized by the extremely randomized trees (ERT) method (figure 2(d)) (Geurts, Ernst and Wehenkel 2006). ERT consist of many tree-based classifiers and are able to classify quickly. The parameters of ERT were set as follows based on the recommended values (Geurts *et al* 2006). Number of trees = 100, number of features randomly selected at each node = $\sqrt{d}$, number of thresholds set by the selected features = 2, where d is the number of dimensions of the feature vector, which is the size of the subimage. The learned ERT ($\mathbf{\Phi}(\cdot)$) outputs the tumor likelihood l of the input feature vector $\mathbf{g} = (g_1, g_2 \cdots, g_d)$.

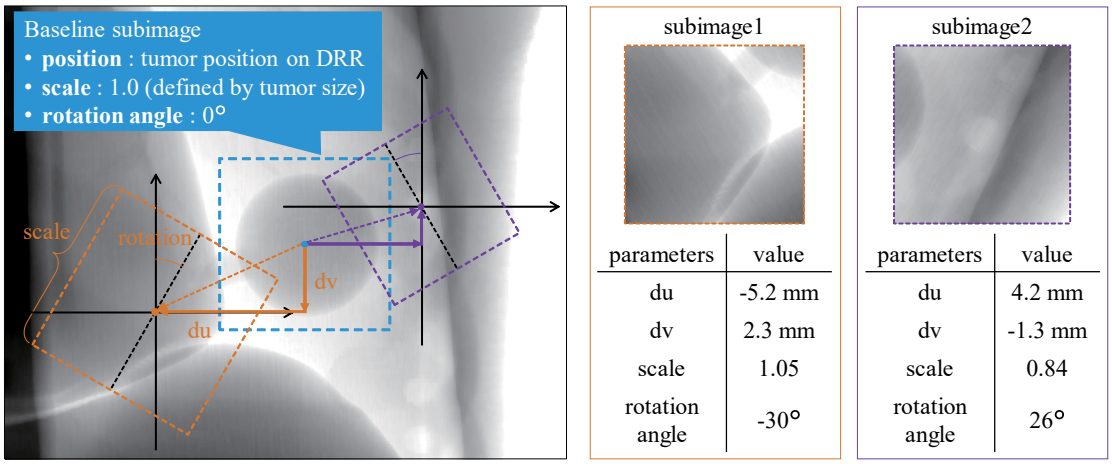

Figure 3. An example of subimage generation. The cropped subimages were created by transforming the baseline subimage with the position (*du*, *dv*), scale and rotation angle parameters. The position and scale of the baseline subimage were determined by the projected tumor position and CTV size on the DRR. Subimages were resized by cropping.

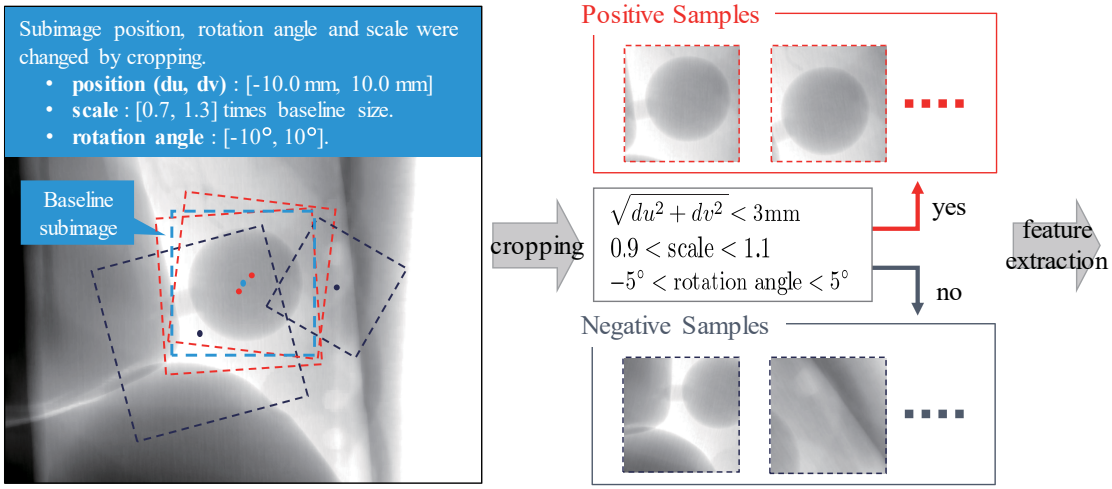

Figure 4. Subimages were cropped while randomly varying the parameters' position, scale, and rotation angle from the baseline subimage. Positive samples were therefore derived from images sampled with a narrow margin around the tumor position (position & 3 mm, scale 0.9–1.1, rotation angle & 5°), and negative samples were made with subimages which differed considerably from the tumor position and contained little or no tumor.

$$\Phi(\mathbf{g}) = l \tag{3}$$

Regression can be used to estimate the tumor position directly from **g**, but regression is generally more difficult than classification. For real-time tumor tracking, we chose the proposed method to solve simpler problems.

2.2.2. Step 2: Likelihood map estimation

Fluoroscopic images were acquired during treatment. A subimage was cropped by changing the incoming fluoroscopic image region, and the machine learning algorithm generated a tumor likelihood map. The tumor likelihood map $L(\mathbf{u})$ was generated following equation.

$$L(\mathbf{u}) = \Phi(\mathbf{g}_\mathbf{u}) \tag{4}$$

First, a subimage at **u** was cropped from the input fluoroscopic image and converted into a feature vector $\mathbf{g}_\mathbf{u}$ by equations (1) and (2). Next, the feature vector $\mathbf{g}_\mathbf{u}$ was input to the ERT, and the tumor likelihood at **u** was calculated. These operations were performed on each pixel position in the search range to obtain $L(\mathbf{u})$ (figure 6). This process was performed with the data from each DFPD, resulting in two likelihood maps ($L_1(\mathbf{u}_1)$ and $L_2(\mathbf{u}_2)$).

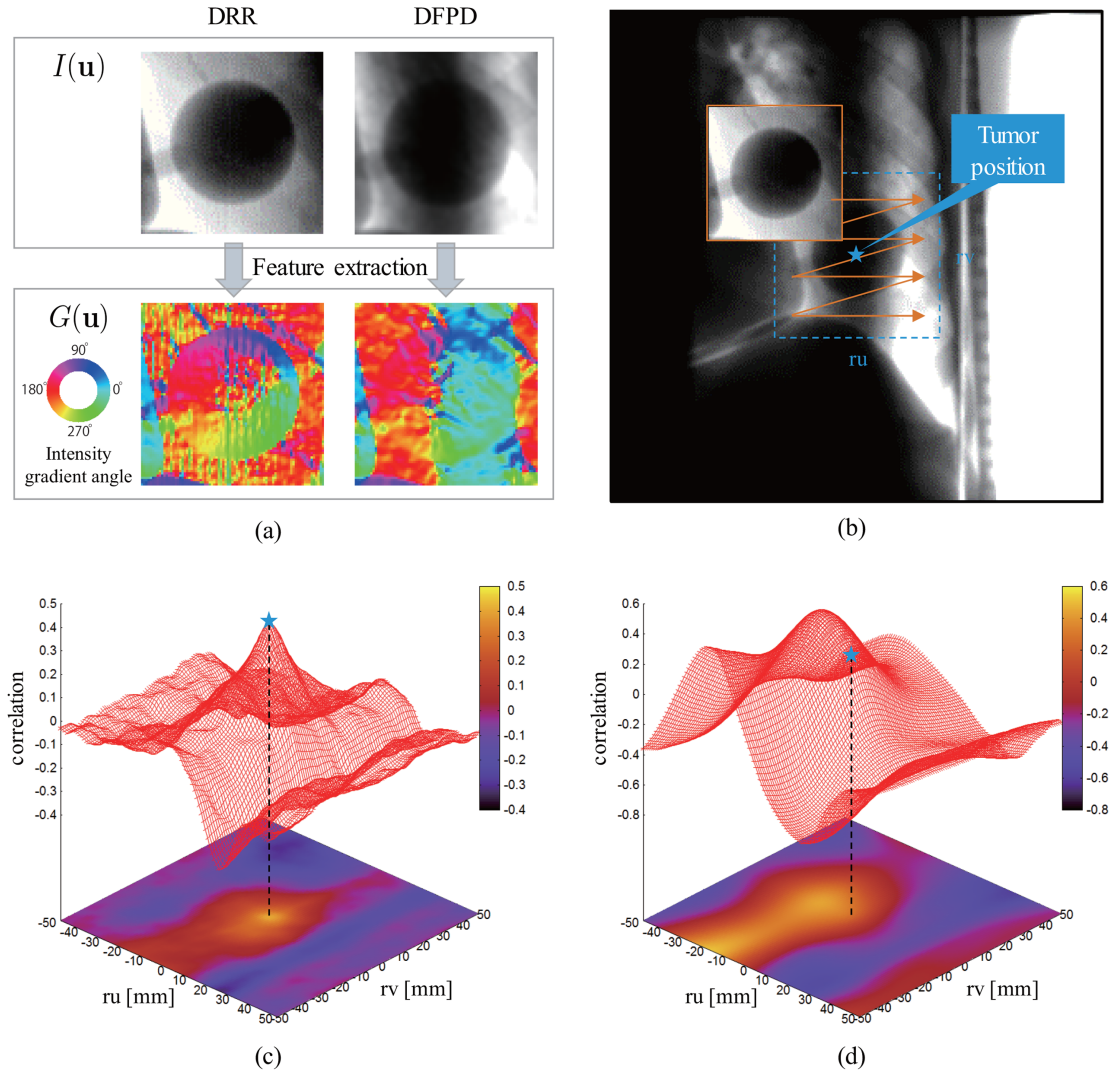

Figure 5. (a) The intensity direction gradient around the tumor position in the DRR and fluoroscopic images. (b) The range where the correlation was calculated between the DRR subimage and the fluoroscopic image. The position of ⋆ was the tumor position on the fluoroscopic image. (c) The correlation of the intensity gradient direction between the DRR subimage and the fluoroscopic image at each position on the fluoroscopic image. The correlation was maximum at the tumor position (position of ⋆). (d) The correlation of the intensity between the DRR subimage and the fluoroscopic image at each position on the fluoroscopic image. The correlation was maximum at a position slightly deviated from the tumor position.

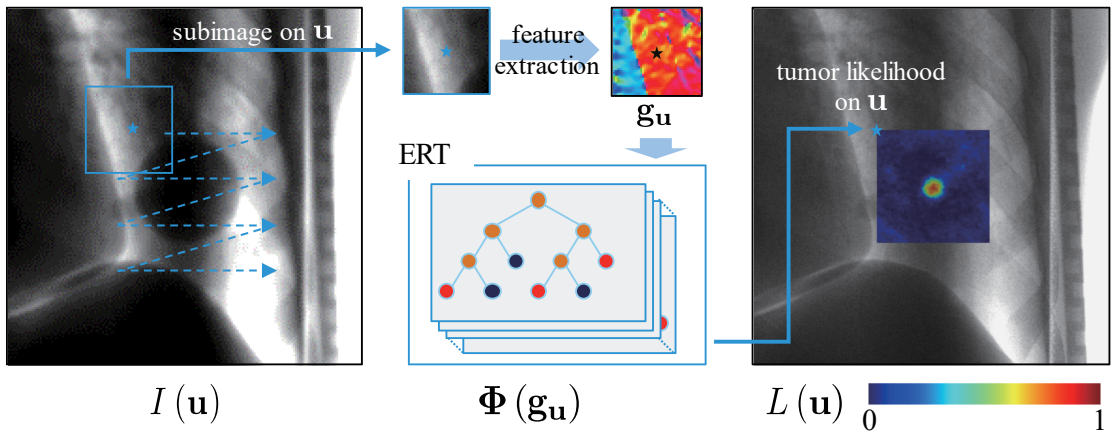

Figure 6. The subimage cropped from the input image was converted into a feature vector, and the tumor likelihood was calculated by ERT.

2.2.3. Step 3: Tumor tracking

Step 3 estimated tumor position in 3D space $\mathbf{X}_t$ at the current frame by using both likelihood maps and prior probability tumor positions in previous frames as follows:

First, to estimate tumor position in 3D space $\mathbf{X}_t$, a likelihood map in 3D space $L_{\text{img}}(\mathbf{X})$ was calculated using equation (5).

$$L_{\text{img}}(\mathbf{X}) = L_1(\mathbf{u}_1)L_2(\mathbf{u}_2) \tag{5}$$

where $\mathbf{u}_1$ and $\mathbf{u}_2$ are projections of the tumor position in 3D $\mathbf{X}$ on each flat panel detector, and they were calculated using equation (6).

$$\lambda_1 \begin{bmatrix} \mathbf{u}_1 \\ 1 \end{bmatrix} = \mathbf{P}_1 \begin{bmatrix} \mathbf{X} \\ 1 \end{bmatrix}, \lambda_2 \begin{bmatrix} \mathbf{u}_2 \\ 1 \end{bmatrix} = \mathbf{P}_2 \begin{bmatrix} \mathbf{X} \\ 1 \end{bmatrix} \tag{6}$$

where $\mathbf{P}_1$ and $\mathbf{P}_2$ are projection matrices of the respective flat panel detector, and λ is a scaling factor. To shorten the calculation time, we did not calculate $L_{\text{img}}(\mathbf{X})$ in every calculation of $\mathbf{X}$, but reused $L_1(\mathbf{u}_1)$ and $L_2(\mathbf{u}_2)$.

Second, tumor positional probability density function $L_{\text{move}}(\mathbf{X})$ was estimated by tumor positions in the previous two frames:

$$L_{\text{move}}(\mathbf{X}) = \frac{1}{\sqrt{(2\pi)^3 |\mathbf{\Sigma}|}} \exp\left(-\frac{1}{2}(\mathbf{X}-\boldsymbol{\mu})\mathbf{\Sigma}^{-1}(\mathbf{X}-\boldsymbol{\mu})^t\right) \tag{7}$$

$$\mathbf{\Sigma} = \alpha\mathbf{I}, \boldsymbol{\mu} = \mathbf{X}_{t-1} + (\mathbf{X}_{t-1} - \mathbf{X}_{t-2}) \tag{8}$$

where $\mathbf{X}_{t-1}$ and $\mathbf{X}_{t-2}$ are tumor position on the previous 1 and 2 frames, respectively, $\mathbf{\Sigma}$ is the covariance matrix of the Gaussian distribution, $\mathbf{I}$ is the 3×3 identity matrix, and α is the scale of the variance. Since the fluoroscopic imaging rate was 15 fps, we assumed that tumor speed in the previous two frames could be roughly the same (Teo, Guo, Ahmed, Alayoubi, Kehler, Fontaine, Sasaki and Pistorius 2019). Therefore, for real-time processing, we used an isotropic Gaussian distribution, which is easier to calculate, as the prior distribution of the tumor motion. The parameter α affects the probability distribution, with a smaller value providing a sharper probability distribution and a higher value, making a flatter curve.

Next, the 3D tumor position was estimated from $L_{\text{img}}(\mathbf{X})$ and $L_{\text{move}}(\mathbf{X})$. The posterior distribution of the tumor position $p(\mathbf{X}|L_{\text{img}})$ with the prior distribution $L_{\text{move}}(\mathbf{X})$ after the tumor likelihood map $L_{\text{img}}(\mathbf{X})$ was observed was calculated from Bayes rule as follows.

$$\begin{aligned} p(\mathbf{X}|L_{\text{img}}) &= \frac{p(L_{\text{img}}|\mathbf{X})p(\mathbf{X})}{p(L_{\text{img}})} \\ &= cL_{\text{img}}(\mathbf{X})L_{\text{move}}(\mathbf{X}) \\ &\propto L_{\text{img}}(\mathbf{X})L_{\text{move}}(\mathbf{X}) \end{aligned} \tag{9}$$

Where c is the normalization constant for $\int p(\mathbf{X}|L_{\text{img}})d\mathbf{X} = 1$. In this paper, the tumor position was estimated as the expectation of the posterior distribution. The tumor position $\mathbf{X}_t$ was calculated as follows.

$$\begin{aligned} \mathbf{X}_t &= \int_R \mathbf{X}p(\mathbf{X}|L_{\text{img}})d\mathbf{X} \\ &= \frac{1}{\sum_{\mathbf{X}\in R} L(\mathbf{X})} \sum_{\mathbf{X}\in R} \mathbf{X}L(\mathbf{X}) \end{aligned} \tag{10}$$

where, R is the search range in 3D space and $L(\mathbf{X}) = L_{\text{img}}(\mathbf{X})L_{\text{move}}(\mathbf{X})$.

The proposed method tracked the tumor by learning the image pattern around the tumor. If the image pattern around the tumor is clearly contained in the fluoroscopic image, it is possible to track the tumor without using the prior distribution by tumor motion. However, the position of the detected tumor may become unstable due to image quality degradation due to noise and estimation error. In order to stabilize the result, the proposed method utilized the prior distribution by tumor motion.

Table 1. Tumor tracking accuracy expressed as Euclidean distance and computation time.

	Tumor position error (mm)			Tracking computation time (ms/frame)
Case number	Mean ± SD	95th percentile	Training time (s)	Mean ± SD
Case 1	1.55 ± 0.77	2.94	81.75	29.57 ± 2.01
Case 2	0.64 ± 0.17	0.96	53.60	28.73 ± 2.00
Case 3	1.53 ± 0.56	2.54	79.60	28.90 ± 1.75
Case 4	0.59 ± 0.44	1.55	48.90	28.30 ± 1.95
Case 5	1.02 ± 0.31	1.48	55.00	28.29 ± 1.83
Case 6	0.52 ± 0.27	0.83	64.95	28.68 ± 1.95
Case 7	1.37 ± 0.54	2.32	70.71	28.46 ± 1.74
Case 8	1.08 ± 0.20	1.44	71.55	28.31 ± 1.85
mean	1.03 ± 0.34	1.76 ± 0.71	65.76 ± 11.48	28.66 ± 1.89

Abbreviation: SD = standard deviation

3. Evaluation

3.1. Tracking accuracy

Retrospective tracking positional analysis was performed on the data of eight patients with lung cancer, who were receiving carbon-ion pencil beam scanning treatment in our hospital. All gave informed consent for use of their data in this study, which was reviewed and approved by the Institutional Review Board of our hospital.

A certified oncologist delineated the gross tumor volume (GTV) on the planning 4DCT at T50. The GTV contour at T50 was transferred to other respiratory phases by deformable image registration (Vincent, Larochelle, Lajoie, Bengio and Manzagol 2010). The tumor contours were then projected onto the DRR images, which were transformed into 2D images, in their respective respiratory phases (4D-DRR). The oncologist and medical physicist independently entered the reference tumor positions (ground truth [GT]) on DFPD images (approximately five respiratory cycles [= 300 frames]) and checked the results together on the basis of the 4D-DRR.

Tracking accuracy was evaluated as the Euclidean distance between the reference and calculated tumor positions.

3.2. Computation time

Computation times for the training (step 1) and tracking (steps 2–3) stages were averaged over all frames and 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). Calculations were performed by multithreading computation (7 cores).

4. Result

4.1. Tracking accuracy

The tumor tracking accuracy averaged over all patients was 1.03 ± 0.34 mm (mean ± SD) and 1.76 ± 0.71 mm (95th percentile) in Euclidean distance (table 1).

Figures 7 and 8 show the best (case 6) and worst (case 1) examples of tumor tracking by the proposed method, respectively. According to the best case, tracking accuracy in the x, y, and z directions was 0.04 ± 0.32 mm, 0.14 ± 0.21 mm and -0.34 ± 0.26 mm, respectively. In the worst case these values were 0.89 ± 0.88 mm, −0.63 ± 0.66 mm, and 0.13 ± 0.75 mm, respectively.

4.2. Computation time

Training time averaged over all patients was 65.76 ± 11.48 s (table 1). Computation time for tumor tracking with two DFPDs averaged over all cases was 28.66 ± 1.89 ms per frame. Since the almost same subimage size was used for all cases, the computation times described above were closely similar for all cases.

5. Discussion

In this study, we evaluated the positional accuracy and computation time of our machine learning-based markerless tracking algorithm using eight lung cancer cases. Our algorithm achieved good tracking

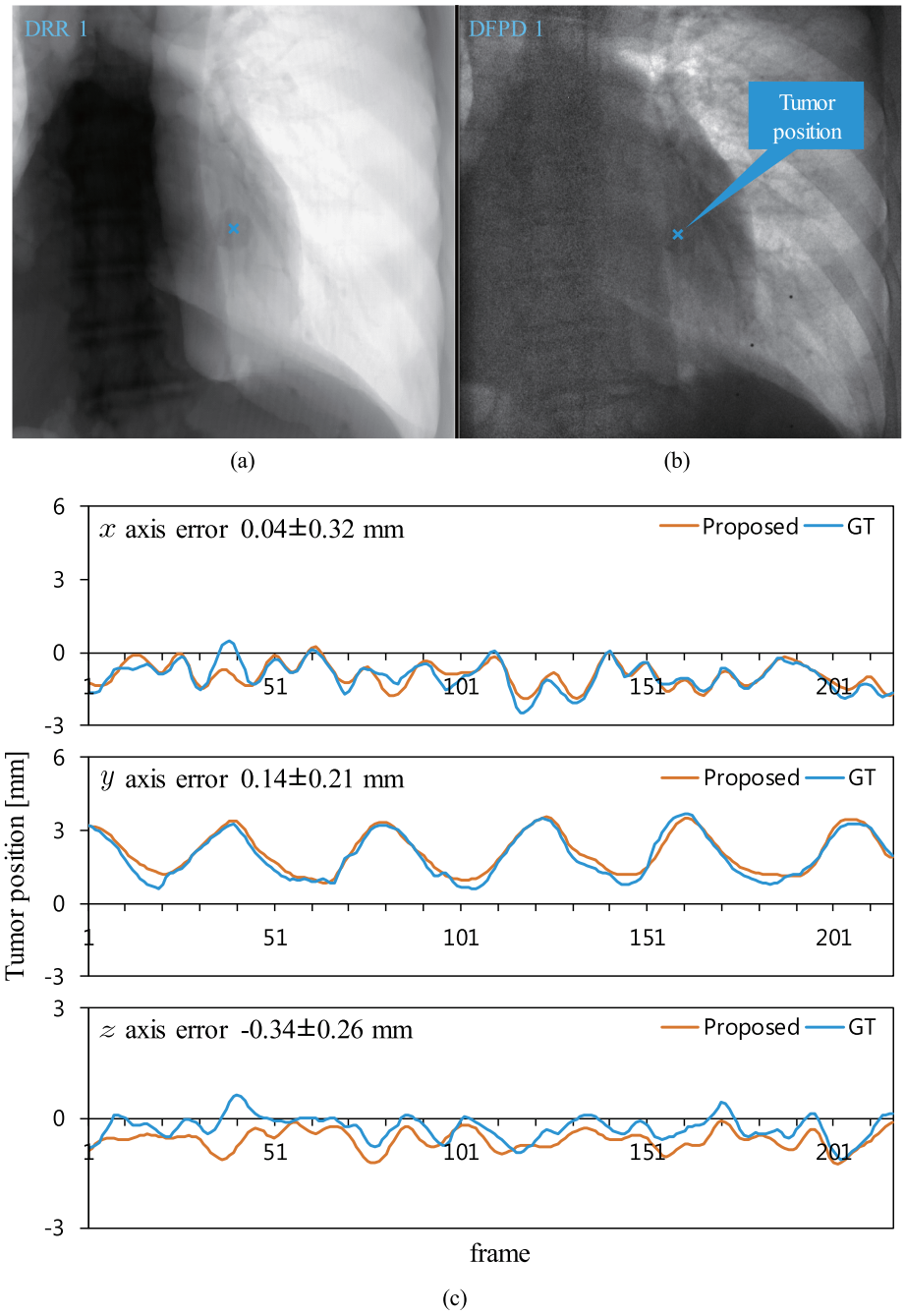

Figure 7. An example of successful tracking (case 6). (a) DRR images and (b) fluoroscopic image on DFPD1. (c) Tumor tracking results in the *x* direction (upper panel), *y* direction (middle panel) and *z* direction (lower panel). *Abbreviations*: DRR = digitally reconstructed radiography, DFPD = dynamic flat panel detector, GT = ground truth.

positional accuracy (= 1.76 ± 0.71 mm) in real time (= 28.66 ± 1.89 ms/frame), suggesting it can be used at a 30-fps acquisition rate.

5.1. Tracking accuracy

An example of a successful case (case 4) shows tumor shape clearly on DRR and DFPD images (figures 9(a) and (b)). Pulmonary vessels are visible on both images; however, image contrast of ribs was relatively low on DRR images compared to DFPD images. The tumor was located in the upper lobe. Tumor displacement in the *y* direction was approximately 5 mm, and the calculated tracking positions were close to the GT positions (figure 9(c)) (0.59 ± 0.44 mm).

Another example (figures 7(a) and (b)) shows that the tumor shape is visualized clearly but overlaid by the heart (case 6). Image contrast of the tumor was slightly lower than that of case 4. Since the proposed method learns the image pattern, it is difficult to track the tumor if the contrast of the image is low. However, in case 6, although image contrast of the tumor is low, the pattern of the pulmonary vessels is characteristic,

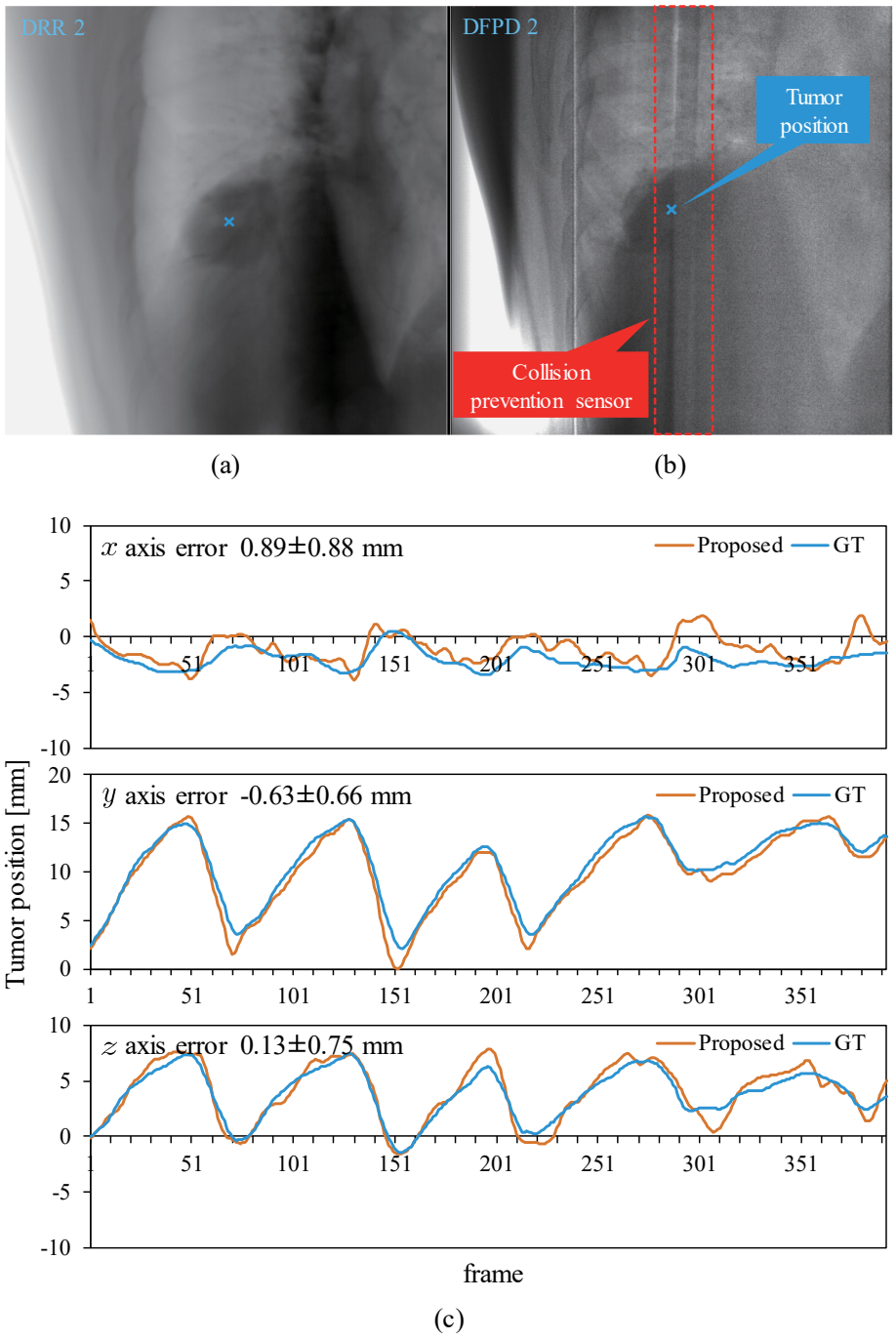

Figure 8. An example of a less successful tracking case (case 1). (a) DRR images and (b) fluoroscopic image on DFPD2. (c) Tumor tracking results in the *x* direction (upper panel), *y* direction (middle panel) and *z* direction (lower panel). *Abbreviations*: DRR = digitally reconstructed radiography, DFPD = dynamic flat panel detector, GT = ground truth.

and the tumor can be tracked using the pulmonary vessel pattern as a clue. Thus, even if the image contrast of a tumor is low, the proposed method can track the tumor with high accuracy if the anatomic pattern around the tumor is characteristic. In particular, the lung has characteristic vessels, hence tracking lung tumor is compatible with the proposed method. The calculated tumor positions in the *x* and *z* directions were affected by cardiac motion; however, tracking accuracy was 0.52 ± 0.31 mm. If the training data contain variations in cardiac motion, the motion pattern of the heart can also be learned, and hence the effect of the cardiac motion on the tracking result is small.

The tracking accuracy of case 1 was degraded due to the collision prevention sensor on the couch (= 1.55 ± 0.77 mm). The sensor was not visualized on DRR images (figure 8(a)) but was on DFPD images (figure 8(b)). Moreover, since the tumor was located close to the sensor and the image contrast of the sensor was higher than that of the tumor, the calculated tumor position was shifted approximately 1 mm in the *x*

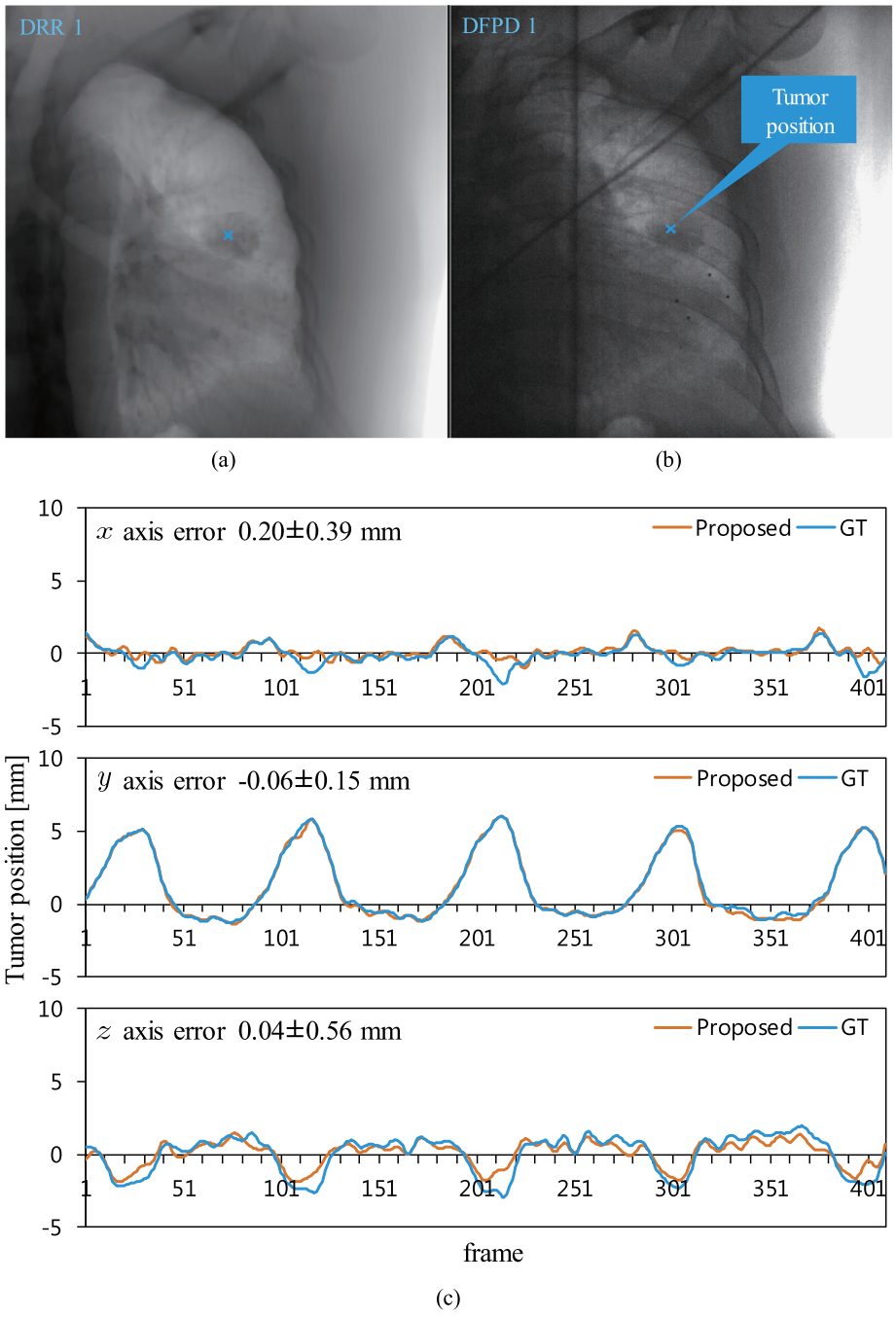

Figure 9. An example of successful tracking (case 4). (a) DRR images and (b) fluoroscopic image on DFPD1. (c) Tumor tracking results in the *x* direction (upper panel), *y* direction (middle panel) and *z* direction (lower panel). *Abbreviations*: DRR = digitally reconstructed radiography, DFPD = dynamic flat panel detector, GT = ground truth.

direction (figure 8(c)). On the other hand, objects with high contrast include the spine. If an object is inside the body, such as the spine, it can be learned from the treatment planning CT because it is contained within the CT. These objects are therefore considered to have little effect on tracking accuracy.

5.2. Computation time

Computation time was strongly dependent on the subimage size. In this study using a subimage size of 60 mm × 60 mm – 90 mm × 80 mm, computation time was 28.66 ± 1.89 ms/frame. However, subimage size could be larger to cover the whole target volume; as a result, the computation time would likely increase. To improve computation time, another solution would be to set subimage position on a part of the target volume.

5.3. Other markerless tracking methods

Reports of other markerless tumor tracking algorithms require tumor position on training fluoroscopic images (Cui *et al* 2007a, 2007b, Zhang *et al* 2014, 2015, Mori *et al* 2016). This manual input of tumor position on fluoroscopic images is time-consuming. Since our algorithm uses 4DCT and tumor position input from the treatment planning system, it does not require manual input of tumor position, obviating the need for extra work.

An automatic markerless tumor tracking algorithm (Shieh *et al* 2017) using treatment planning CT had an accuracy of 1.6–2.9 mm in 3D Euclidean distance, which was inferior to our result (0.52–1.6 mm).

Deep neural networks (DNN) represent a relatively new technique and have improved performance in artificial intelligence more than conventional machine learning (Vincent *et al* 2010, Masci, Meier, Cireşan and Schmidhuber 2011). Our group previously published a DNN-based markerless tumor tracking algorithm and evaluated tracking positional accuracy using lung and liver cases (Hirai, Sakata, Tanizawa and Mori 2019). Although the training time for DNN required several hours, positional accuracy and computation time were 2.25 $\pm$ 1.03 mm and 39.9 $\pm$ 3.4 ms for lung cases, respectively. While our algorithm is based on conventional machine learning, its tracking accuracy and computation time are superior to those of the DNN algorithm. The model was designed for individual patients, and therefore does not require greater robustness. As DNN requires very large training datasets to optimize large numbers of parameters, it is difficult to prepare very large training datasets from one patient. Our markerless tumor tracking was achieved by the conventional machine learning approach, which does not require such large datasets.

5.4. Treatment throughput

Currently, our institute has started respiratory-gated carbon-ion scanning beam treatment using markerless tracking (Mori *et al* 2016). It requires template images, which are generated in the preparation stage just after the patient setup procedure. In our clinical experience, this takes an average of 8.6 min. Use of our new algorithm could improve treatment throughput by omitting this template image preparation process. Since we perform markerless tracking in the fixed-beam-port treatment room (figure 1), once a template image is prepared, it can be reused when horizontal and orthogonal beams are used in the same fraction. However, in the rotating gantry treatment room, the DFPDs change position as a function of gantry angle. Therefore, a separate template image must be prepared for each gantry angle. This may slow throughput. Use of our algorithm could solve this problem.

5.5. Study limitations

Several limitations of this study warrant mention. First, a certified oncologist and medical physicist input the reference tumor position by comparing 4D-DRR images with the target counters, but this does not guarantee the elimination of observer error. However, we carefully checked the reference tumor position to decrease this error.

Second, our algorithm considered organ deformation only on 4D-DRR images. However, we applied the data augmentation process in translation and rotation but not deformation on the training data to prevent the degradation of tracking accuracy due to interfractional changes. Even though our algorithm calculated tumor position using a classification model, it might still have oversimplified the deformation. Therefore, if the machine learning approach had not worked well, we would have been unable to continue treatment because it is impossible to modify the model to estimate tumor positions with accuracy. One solution is to retrain the model (fine-tuning) using additional 4DCT images acquired before each fraction. This limitation will apply to any study using the machine learning approach to calculate a model using planning CT data.

Third, we did not evaluate our algorithm for abdominal organs such as liver and pancreas. This is because it is still difficult to capture tumor position due to low contrast. At the present time, use of external gating or insertion of implanted fiducial markers are still standard practice. In this paper, we utilized an isotropic Gaussian distribution as the prior distribution of the tumor motion. However, if the covariance matrix of the tumor motion is able to be estimated from the 4DCT, the tracking results may be more stable. It may help in tracking the tumor from low contrast images. We hope our new method will represent an improvement on these.

6. Conclusions

We developed a machine learning-based markerless tumor tracking algorithm. Our algorithm trained on the tumor image patterns from the DRR images generated from the treatment planning CTs and tracked the tumors in the fluoroscopic images fully automatically. By learning the tumor image patterns by ERT using the intensity gradient direction as the feature descriptor to reduce the image quality differences between the DRR and the fluoroscopic images, markerless tumor tracking can be performed with high accuracy and in

real time. Allowing that the evaluation was conducted in a limited number of lung cases, the algorithm achieved good tracking accuracy in a short computational time. We consider that our algorithm will be highly useful in improving gate treatment accuracy.

Acknowledgment

Dr Sakata, Mr Hirai, Ms Kobuna and Mr Tanizawa are employed by Toshiba Corp., Kawasaki, Japan. We thank Libby Cone, MD, MA, from DMC Corp. (www.dmed.co.jp) for editing drafts of this manuscript.

References

Bryant J, Rottmann J, Lewis J, Mishra P, Keall P and Berbeco R 2014 Registration of clinical volumes to beams-eye-view images for real-time tracking *Med. Phys.* **41** 121703

Cui Y, Dy J G, Sharp G C, Alexander B and Jiang S B 2007a Multiple template-based fluoroscopic tracking of lung tumor mass without implanted fiducial markers *Phys. Med. Biol.* **52** 6229–42

Cui Y, Dy J G, Sharp G C, Alexander B and Jiang S B 2007b Robust fluoroscopic respiratory gating for lung cancer radiotherapy without implanted fiducial markers *Phys. Med. Biol.* **52** 741–55

Geurts P, Ernst D and Wehenkel L 2006 Extremely randomized trees *Mach. Learn.* **63** 3–42

Hirai R, Sakata Y, Tanizawa A and Mori S 2019 Real-time tumor tracking using fluoroscopic imaging with deep neural network analysis *Phys. Medica* **59** 22–9

Jakel O, Karger C and Debus J 2009 The future of heavy ion radiotherapy *Med. Phys.* **35** 5653–63

Kubo H D and Hill B C 1996 Respiration gated radiotherapy treatment: a technical study *Phys. Med. Biol.* **41** 83–91

Li R, Lewis J, Cervino L and Jiang S 2009 A feasibility study of markerless fluoroscopic gating for lung cancer radiotherapy using 4DCT templates *Phys. Med. Biol.* **54** N489–500

Masci J, Meier U, Cireşan D and Schmidhuber J 2011 Stacked convolutional auto-encoders for hierarchical feature extraction eds T Honkela, W Duch, M Girolami and S Kaski *Artificial Neural Networks and Machine Learning – ICANN 2011* (Berlin: Springer) pp 52–9

Meschini G, Seregni M, Pella A, Ciocca M, Fossati P, Valvo F, Riboldi M and Baroni G 2017 Evaluation of residual abdominal tumour motion in carbon ion gated treatments through respiratory motion modelling *Phys. Medica* **34** 28–37

Mori S *et al* 2012 First clinical experience in carbon ion scanning beam therapy: retrospective analysis of patient positional accuracy *J. Radiat. Res.* **53** 760–8

Mori S *et al* 2016 Carbon-ion pencil beam scanning treatment with gated markerless tumor tracking: An analysis of positional accuracy *Int. J. Radiat. Oncol. Biol. Phys.* **95** 258–66

Mori S, Karube M, Yasuda S, Yamamoto N, Tsuji H and Kamada T 2017 Gating window dependency on scanned carbon-ion beam dose distribution and imaging dose for thoracoabdominal treatment *British J. Radiol.* **90** 20160936

Mori S, Shirai T, Takei Y, Furukawa T, Inaniwa T, Matsuzaki Y, Kumagai M, Murakami T and Noda K 2012 Patient handling system for carbon ion beam scanning therapy *J. Appl. Clin. Med. Phys./Am. Coll. Med. Phys.* **13** 3926–40

Rottmann J, Aristophanous M, Chen A, Court L and Berbeco R 2010 A multi-region algorithm for markerless beam's-eye view lung tumor tracking *Phys. Med. Biol.* **55** 5585–98

Sharp G, Kandasamy N, Singh H and Folkert M 2007 GPU-based streaming architectures for fast cone-beam CT image reconstruction and demons deformable registration *Phys. Med. Biol.* **52** 5771–83

Shieh C-C A, Caillet V, Dunbar M, Keall P, Booth J, Hardcastle N, Haddad C, Eade T and Feain I 2017 A Bayesian approach for three-dimensional markerless tumor tracking using kV imaging during lung radiotherapy *Phys. Med. Biol.* **62** 3065–80

Shieh C-C A, Keall P, Kuncic Z, Huang C-Y and Feain I 2015 Markerless tumor tracking using short kilovoltage imaging arcs for lung image-guided radiotherapy *Phys. Med. Biol.* **60** 9437–54

Shirato H *et al* 2003 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.* **56** 240–7

Teo P, Crow R, Van Nest S, Sasaki D and Pistorius S 2013 Tracking lung tumour motion using a dynamically weighted optical flow algorithm and electronic portal imaging device *Meas. Sci. Technol.* **24** 074012

Teo P, Guo K, Ahmed B, Alayoubi N, Kehler K, Fontaine G, Sasaki D and Pistorius S 2019 Evaluating a potential technique with local optical flow vectors for automatic organ-at-risk (OAR) intrusion detection and avoidance during radiotherapy *Phys. Med. Biol.* **64** 145008

Vincent P, Larochelle H, Lajoie I, Bengio Y and Manzagol P-A 2010 Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion *J. Mach. Learn. Res.* **11** 3371–408

Zhang X, Homma N, Ichiji K, Abe M, Sugita N, Takai Y, Narita Y and Yoshizawa M 2014 A kernel-based method for markerless tumor tracking in kV fluoroscopic images *Phys. Med. Biol.* **59** 4897–911

Zhang X, Homma N, Ichiji K, Takai Y and Yoshizawa M 2015 Tracking tumor boundary in MV-EPID images without implanted markers: A feasibility study *Med. Phys.* **42** 2510–23