Radiotherapy and Oncology 138 (2019) 25–29

Contents lists available at ScienceDirect

Radiotherapy and Oncology

journal homepage: www.thegreenjournal.com

Technical Note

Multi-object tracking in MRI-guided radiotherapy using the tracking-learning-detection framework

Jennifer Dhont [a,b,c,*], Jef Vandemeulebroucke [a,b], Davide Cusumano [d], Luca Boldrini [d], Francesco Cellini [d], Vincenzo Valentini [d,e], Dirk Verellen [c,f]

[a] Vrije Universiteit Brussel (VUB), Department of Electronics and Informatics (ETRO); [b] Imec, Leuven; [c] Vrije Universiteit Brussel (VUB), Faculty of Medicine and Pharmacy, Belgium; [d] Fondazione Policlinico Universitario "A. Gemelli" IRCCS; [e] Università Cattolica del Sacro Cuore, Roma, Italy; [f] Department of Radiotherapy, GZA Ziekenhuizen – Sint Augustinus, Iridium Kankernetwerk, Antwerp, Belgium

ARTICLE INFO

Article history:
Received 5 December 2018
Received in revised form 3 May 2019
Accepted 3 May 2019
Available online 25 May 2019

Keywords:
Motion management
MRI
Tracking

ABSTRACT

The application of the tracking-learning-detection (TLD) framework; a performant tracking algorithm for real-life objects in CCD video, was evaluated and successfully optimized for tracking anatomical structures in low-quality 2D cine-MRI acquired during MRI-guided radiotherapy. Sub-pixel tracking accuracy and >95% precision and recall was achieved despite significant deformations and periodical disappearance.

Technological progress over the past decade has led to the commercialization of MRI-guided radiotherapy (MRIgRT), bringing the epitome of image-guided RT (IGRT) into clinical practice [1,2]. Before radiation delivery, the anatomy of the day and inter-fraction variations can be visualized by high soft-tissue contrast on a 3D MRI, while no additional imaging dose is delivered to the patient. When large variations with the planned scenario near target or organs-at-risk (OARs) are observed, dose delivery can be adapted with the rendering of a synthetic CT, creating a plan-of-the-day [3,4].

With 2D cine-MRI during delivery, patient anatomy and intra-fraction motion can be monitored directly and in real-time, again minimizing the geometrical uncertainty with no additional imaging dose [5]. Especially for lesions subject to respiratory-induced motion, the availability of cine-MRI during gating or tracking could be of great benefit as no implanted fiducials and/or the use of an external respiratory signal are required for target localization [6,7].

Based on average human breathing and reported respiratory-induced motion ranges, an adequate cine-MRI frame rate for monitoring respiratory motion is estimated to be around 2 Hz or more [8,9]. At this frequency, cine-MRI is only available on low-field MRIgRT systems and is limited to a single 2D imaging plane [10–12]. Consequently, as respiratory motion follows a 3D trajectory, with the presence of hysteresis and baseline drift, out of plane motion of the imaged structures is likely [13]. If so, the appearance will change continuously on the cine-MRI image, with complete dis- and re-appearance in case of high amplitude motion perpendicular to the imaging plane. The relatively high frequency needed for respiratory motion monitoring also poses limitations on the spatial resolution and signal-to-noise ratio (SNR), the latter already compromised by the low magnetic field strength [14].

As such, accurately tracking anatomical structures on cine-MRI during RT is not straightforward. Published and available tracking algorithms for cine-MRI are often validated on structures relatively stable in appearance, and images with spatial resolution an order of magnitude higher [15–18].

In computer vision, a tracking-learning-detection (TLD) framework proposed by Kalal et al. [19] was shown to be a highly performant approach to object tracking, demonstrated to outperform multiple other tracking algorithms on video streams of moving pedestrians, cars, and other real-life objects. It allows for real-time multi-object tracking and is able to cope with large variations in the object appearance over time, including complete dis- and re-appearance, making it an ideal candidate to tackle some of the limitations presented above. However, the TLD framework has only been applied and validated on non-medical CCD camera video streams, with high SNR, high spatial resolution and feature rich objects.

In this study TLD was validated as a tracking algorithm for anatomical structures on low-field cine-MRI, and optimized to increase tracking accuracy, using real patient cine-MRI sequences acquired during RT.

* Corresponding author at: Department of Electronics and Informatics (ETRO), Vrije Universiteit Brussel (VUB), Pleinlaan 2, B-1050 Brussels, Belgium.
E-mail address: jennifer.dhont@vub.be (J. Dhont).

https://doi.org/10.1016/j.radonc.2019.05.008
0167-8140/

Materials and methods

Original TLD

What follows is a brief overview of the original TLD framework, and how it can be incorporated into the cine-MRI during RT workflow. A detailed description of the TLD framework and its performance compared to other tracking algorithms outside the medical field can be found elsewhere [19].

TLD consists of a tracker for frame-to-frame motion estimation, an object detector, and a learning framework. The tracker is based on Median-Flow using a pyramidal Lucas-Kanade tracker, extended with failure detection [20,21]. The object detector is a cascaded classifier consisting of patch variance analysis, an ensemble classifier based on independent pixel comparisons, and a nearest neighbor classifier.

The workflow of TLD and more specifically the learning framework can be divided in two parts; an initiation phase at the start of each treatment fraction, and run-time during treatment. During initiation, the object(s) to be tracked is defined on the first cine-MRI frame, either by registration to the planning images or by manual selection of a bounding box by an experienced user. After object definition, 200 possible future appearances of the object(s) are rendered synthetically through geometrical transformations (in-plane rotation, scaling and translations) of the defined object on the first frame. The remainder of the frame is sampled to create negative examples, and the object detector is trained a first time.

During run-time, with each new incoming cine-MRI frame, the tracker and detector independently determine the object location, together with a confidence measure calculated using normalized cross correlation (NCC)-based similarity with previous positive patches [19]. An integrator then combines the tracker and detector locations, outputting the maximally confident object patch in the new frame. If both the tracker and detector output no location, the object is defined as 'not visible' in the current frame. If so, the object detector has to pick up re-appearance of the object in following frames, as frame-to-frame trackers require a previous location.

During run-time and after each frame, the detector output is automatically scrutinized and false-negatives and -positives are estimated based on a most-likely trajectory of the object and its unique location in every frame, respectively. The detector is then re-trained on the corrected data, increasing its generalization and discrimination capabilities, improving the detector accuracy during run-time [22,23].

Multi-object tracking is performed by running the workflow for each object in parallel.

TLD optimization

The following alterations were made to make the original TLD algorithm more appropriate for the cine-MRI application.

During the initiation phase, the geometrical transformations to render synthetic object appearances were changed from ±10° to ±5° (1° step) in-plane rotation, ±1% to ±10% (1% step) scale change and ±1% to no translations, rendering an equal number of synthetic positives. Quantities were chosen based on published motion data of anatomical structures [24].

The tracker implemented in the original TLD framework determines the object motion by estimating the motion of 10×10 points on a grid sampled inside the structure ± 5-pixel margin. Frame-to-frame motion of every point is estimated by locally solving the optical flow equation. This motion estimation approach is known to fail when the area around the point is uniform, and motion is small. As most anatomical structures on cine-MRI data are largely uniform and lack sharp heterogeneous regions, as seen in Fig. 1, the margin was increased to include sufficient object boundary and as such more heterogeneity. A step-wise increase resulted in an optimal ± 10-pixel margin.

Failure detection on the tracker was expanded by putting a limit of 10 mm in each direction on the allowed frame-to-frame motion, based on reported respiratory-induced tumor motion in literature [25,26], and considering the used frame rate. Furthermore, and based on the same considerations, the search-space of the detector was limited to an area of 100×100 mm^2 centered around the previous object location, unless there was no previous object location.

Data and accuracy analysis

Validation of the TLD framework and optimization was carried out retrospectively using 2D cine-MRI data acquired during MRI-guided patient treatment on MRIdian (ViewRay Inc., Cleveland, Ohio, USA), equipped with a 0.35 T MR scanner [10]. During treatment, cine-MRI consists of the acquisition of one user defined sagittal plane of 7 mm thickness, using a true fast imaging with steady state precession (TrueFisp) sequence, with an acquisition

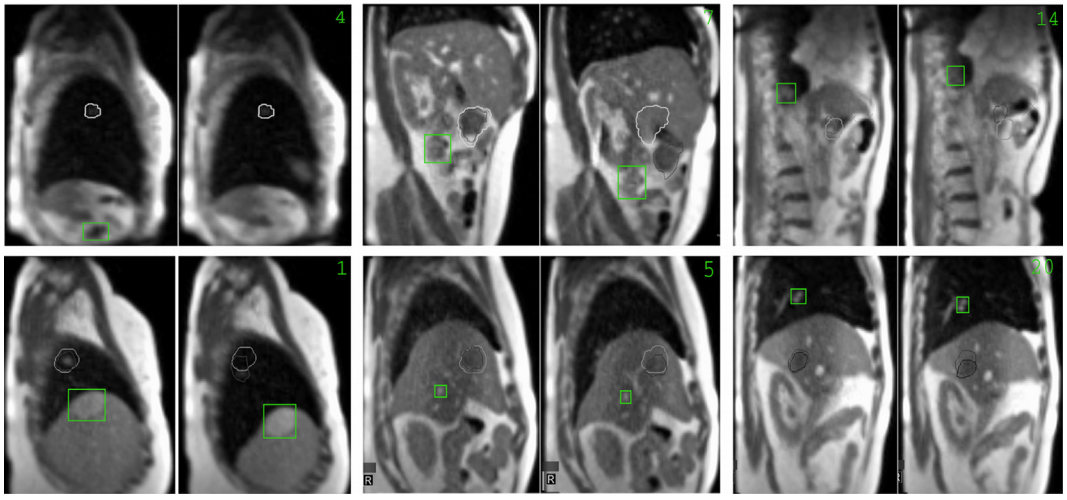

Fig. 1. Illustration of the objects (in rectangle) with highest tracking-error (obj. 4, 7 and 14, above) or low recall (obj. 1, 5 and 20, below) when applying original TLD. Two cine-MRI frames are shown per object; the first frame used for initiation and a later frame to indicate the extent of deformation and motion, or disappearance (obj. 4). Visible embedded treatment field and object contours by the system's software motivate the choice of other anatomical objects for accuracy evaluation.

frequency of 4 frames/second (s) and a spatial resolution of 3.5×3.5 mm^2 [27]. However, due to vendor restrictions, acquired cine-MRI data is only available post-treatment in an upsampled (0.4×0.4 mm^2) OGV video file format, with embedded moving target and fixed treatment field contours, as seen in Fig. 1.

To avoid interference by these contours, 20 non-target objects were chosen for validation (5 liver, 8 abdominal and 7 thoracic structures), ranging from unique and stable in appearance (obj. 8, 11, 12, 17 and 20) to highly deforming (obj. 1, 6, 7, 10 and 13), or very similar to other objects in the frame (obj. 2, 5, 15, 16 and 18); illustrated in Supplementary Fig. 1. Objects were also either static or mobile due to respiration or other physiological processes. Specifically, objects 4, 6, 9 and 10 moved perpendicular to the imaging plane due to respiration, causing continuous changes in appearance with respect to the first frame, with periodic full dis- and re-appearance of object 4. Location and maximum excursion in the AP and CC direction are included per object in supplementary Table 1.

Supplementary Table 1 also includes the number of cine-MRI frames used per object for validation, ranging from 42 to 1000 frames. Tracking accuracy was evaluated by calculating the tracking error (TE) defined as the root-mean-square error (RMSE) between TLD output and ground-truth object center-of-mass (COM). Ground-truth was determined using the scale-invariant feature transform (SIFT) followed by manual validation by an expert [28]. TLD tracking precision and recall were also evaluated per object. Precision was calculated as the frame ratio of true positive detections over positive detections by TLD, while recall was calculated as the frame ratio of true positive detections by TLD over the number of frames on which the object is visible according to the ground truth.

Dependency of TLD accuracy on the bounding box definition during initiation was analyzed as described in Supplementary materials.

Object pairs 2–16, 3–4, 6–7, 8–9, and 11–17 belonged to the same patient and cine-MRI sequence, and were therefore used to validate the performance of multi-object tracking by comparing the single-object and multi-object TE.

Both TLD performance and the accuracy analysis were carried out using MATLAB (Mathworks, Natick, MA, USA), while SIFT was applied using the OpenCV-Python library, on a general-purpose notebook PC.

Results

The duration of the initiation phase was mostly influenced by the time needed for object definition, as synthetic example rendering, background sampling and training took under 2 s on a general notebook PC. Original and optimized TLD performance was feasible for both single and multi-object tracking at a maximum of 5 frames/s on a standard notebook PC implemented in MATLAB, allowing further time optimization when necessary. As described in the Supplementary materials, the dependency of TLD accuracy on the bounding box definition during initiation was below image resolution for both original and optimized TLD.

Fig. 2 displays the tracking accuracy as mean TE ± 1SD, as well as the 95th percentile, in the anterior-posterior (AP) and cranial-caudal (CC) direction, both from original and optimized TLD and for each object separately.

Original TLD TE averaged over all objects ± 1SD equaled 2.1 ± 1.1 mm and 2.3 ± 1.1 mm in the AP and CC direction, respectively. The 95th percentile over all objects ranged from 1.4 to 7.7 mm and 1.4 to 8.9 mm, in the AP and CC direction, respectively. Tracking precision was above 95% for all objects. However, for three objects (obj. 1, 5 and 20), tracking recall was below 95%, at 93% and 48% and 65%, respectively. The objects with the highest TE (obj. 4, 7 and 14) and lowest recall (obj. 1, 5 and 20) and their extent of deformation and displacement are displayed in Fig. 1.

Optimized TLD TE averaged over all objects ± 1SD equaled 1.6 ± 0.6 mm and 1.9 ± 0.7 mm in the AP and CC direction, respectively. The 95th percentile over all objects ranged from 1.5 to 5.8 mm and 1.5 to 6.0 mm, in the AP and CC direction, respectively.

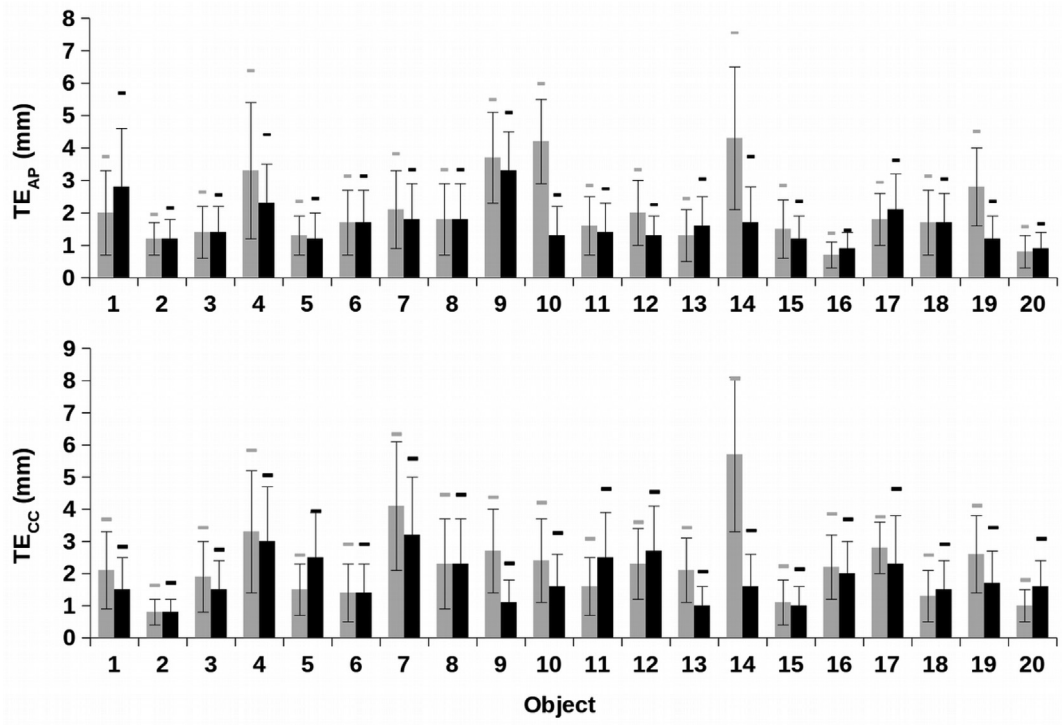

Fig. 2. Tracking errors (TE) illustrated as mean ± 1SD, together with the 95th percentile (horizontal bars), in the anterior-posterior (above) and cranial-caudal (below) direction, both for original (gray) and optimized (black) TLD, for each object separately.

Optimized TLD had a precision above 98% for all objects. Recall was above 95% for all objects, except for object 5 where recall was 70%.

Using the student's t-test, there was no significant difference between the TEs of objects in different locations, i.e. lung, liver or abdomen (AP, CC: $p > 0.05$), nor was the difference between the TEs of original TLD and optimized TLD significantly different (AP: $p = 0.11$, CC: $p = 0.06$).

Multi-object tracking did not change the TE, precision or recall, compared to single-object tracking.

Discussion

In this study, the TLD framework, shown to be a performant tracking approach in computer vision, was introduced to the MRIgRT workflow, and evaluated on a collection of anatomical structures on cine-MRI acquired during MRIgRT. As tracking anatomical structures on cine-MRI frames significantly differs from tracking real-life objects on high-resolution CCD camera video data, the original TLD was adapted in a successful attempt to increase tracking accuracy. Mean TEs of optimized TLD were sub-pixel (<3.5 mm) for every object, including those highly deforming or periodically disappearing from view.

Although optimization was necessary, several benefits of the original TLD framework that made it superior to other tracking algorithms, and promising for the use in the cine-MRI application, remained applicable.

As the detector in TLD is further trained with images acquired during run-time, the TLD approach does not require a pre-treatment training session in an already complicated treatment workflow. In addition, the accuracy of TLD does not depend on how well the objects remain similar to their appearance during initiation, as new appearances are learned over time. Seregni et al. [16] validated a hybrid tracking approach consisting of optical-flow registration, optimized for speed by similarity matching and applying motion fields calculated during an initiation phase. In 64.3% of analyzed trajectories, the tracking error was sub-pixel, but the optimization was only applicable when the object's appearance remained stable. While the accuracy reported is similar, the authors mention out-of-plane motion might cause the tracking error to increase, to which the TLD framework was shown to be robust. Bourque et al. [17] proposed an optimized particle filter-based tracking algorithm for cine-MRI tracking. Similarly, sub-pixel tracking errors were reported, but the authors discuss out-of-plane motion might cause the accuracy to deteriorate, especially in case of dissimilarities with the reference images which have to be acquired and delineated before each treatment fraction. Mazur et al. [18] proposed SIFT as a tracking algorithm and validated on a similar dataset of 0.35 T cine-MRI. Similarly, an optimization of the original SIFT algorithm was performed to increase tracking accuracy, using harmonic analysis. Mean tracking errors were reported to be sub-pixel, but a short training session and knowledge of the algorithm by the user are necessary as suitable parameters for SIFT descriptor construction have to be tweaked on a patient level before tracking.

Although Cerviño et al. [29] showed that a less complex technique based on template-matching and surrogate tracking has superior accuracy over an artificial neural network (ANN), the interest and progress in ANNs over the last few years has increased tremendously in all areas of radiotherapy [30]. As future work, it would be of interest to compare TLD to the current state-of-the-art ANN-based tracking techniques [31,32].

Several tracking algorithms specifically for the cine-MRI in MRIgRT have been proposed in literature [16–18,30–32]. However, only few have been validated on the low-resolution, low SNR images that are acquired with the low-field MRIgRT systems. This large variety in validation datasets complicates an objective comparison of performance between algorithms. All algorithms report sub-pixel tracking accuracy, but resolution in the validation datasets ranges from 1 to 3.5 mm. Variations in magnetic field strength, image protocols, spatial resolution, SNR, the duration of the cine-MRI sequences and evaluated object sites further stresses the need for a publicly available clinical dataset for validation and comparison of feasible tracking algorithms. It should also be noted that tracking irregular objects moving perpendicular to the 2D imaging plane inherently causes tracking errors in the other directions, independent of the tracking algorithm. Further research into multiplane cine-MRI imaging and tracking is therefore of great interest [33,34].

Lastly, in this study the use of TLD on multiple objects was shown to have identical accuracy as single-object tracking. With multi-object tracking, gating could be performed not only based on target-in-field criteria, but also with organs-at-risk (OAR)-out-of-field criteria. Similarly, tumor tracking could be performed by also monitoring nearby OARs, giving information on the non-target structures coming in-field while the tumor and treatment beam move. In case of no respiratory motion, multi-object TLD could be applied to target and OARs simultaneously for general monitoring of possible motion due to other physiological processes.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.radonc.2019.05.008.

References

[1] Lagendijk JJW, Raaymakers BW, Raaijmakers AJE, et al. MRI-linac integration. Radiother Oncol 2008;86:25–9.
[2] Pollard JM, Wen S, Sadagopan R, et al. The future of image-guided radiotherapy will be MR guided. Br J Radiol 2017;90:20160667.
[3] Pathmanathan AU, van As NJ, Kerkmeijer LGW, et al. Magnetic resonance imaging-guided adaptive radiation therapy: a "Game Changer" for Prostate treatment? Int J Radiat Oncol Biol Phys 2018;100:361–73.
[4] Kupelian P, Sonke JJ. Magnetic resonance-guided adaptive radiotherapy: a solution to the future. Semin Radiat Oncol 2014;24:227–32.
[5] Bruijnen T, Stemkens B, Terhaard CHJ, et al. Intrafraction motion quantification and planning target volume margin determination of head-and-neck tumors using cine magnetic resonance imaging. Radiother Oncol 2019;130:82–8.
[6] Henke LE, Contreras JA, Green OL, et al. Magnetic resonance Image-guided radiotherapy (MRIgRT): a 4.5-year clinical experience. Clin Oncol 2018;30:720–7.
[7] Kim JI, Lee H, Wu HG, et al. Development of patient-controlled respiratory gating system based on visual guidance for magnetic resonance image-guided radiation therapy. Med Phys 2017;44:4838–46.
[8] Keall P, Mageras GS, Balter JM, et al. The management of respiratory motion in radiation oncology report of AAPM Task Group 76. Med Phys 2006;33:3874–900.
[9] Shannon CE. Communication in the presence of noise. Proc. IRE 1949;37:10–21 (Reprint as classic paper in: Proceedings of the IEEE. 1998 Feb; 86(2)).
[10] Mutic S, Dempsey JF. The ViewRay system: magnetic resonance-guided and controlled radiotherapy. Semin Radiat Oncol 2014 Jul;24:196–9.
[11] Raaymakers BW, Jurgenliemk-Schulz IM, Bol GH, et al. First patients treated with a 1.5 T MRI-Linac: clinical proof of concept of a high-precision, high-field MRI guided radiotherapy treatment. Phys Med Biol 2017;62:41–50.
[12] Sawant A, Keall P, Pauly KB, et al. Investigating the feasibility of rapid MRI for image-guided motion management in lung cancer radiotherapy. Biomed Res Int 2014;2014:485067.
[13] Seppenwoolde Y, Shirato H, Kitamura K, et al. Precise and real-time measurement of 3D tumor motion in lung due to breathing and heartbeat, measured during radiotherapy. Int J Radiat Oncol Biol Phys 2002;53:822–34.
[14] Maubon AJ, Ferru JM, Berger V, et al. Effect of field strength on MR images: comparison on the same subject at 0.5, 1.0, and 1.5 T. Radiographics 1999;19:1057–67.
[15] Shi X, Diwanji T, Mooney KE, et al. Evaluation of template matching for tumor motion management with cine-MR images in lung cancer patients. Med Phys 2014;41:052304.
[16] Seregni M, Paganelli C, Summers P, et al. A hybrid image registration and matching framework for real-time motion tracking in MRI-guided radiotherapy. IEEE Trans Biomed Eng 2018;65:131–9.
[17] Bourque AE, Bedwani S, Carrier JF, et al. Particle filter-based target tracking algorithm for magnetic resonance-guided respiratory compensation:

robustness and accuracy assessment. Int J Radiat Oncol Biol Phys 2018;100:325–34.
[18] Mazur TR, Fischer-Valuck BW, Wang Y, et al. SIFT-based dense pixel tracking on 0.35 T cine-MR images acquired during image-guided radiation therapy with application to gating optimization. Med Phys 2016;43:279.
[19] Kalal Z, Mikolajczyk K, Matas J. Tracking-learning-detection. IEEE Trans Pattern Anal Mach Intell 2012;34:1409–22.
[20] Kalal Z, Mikolajczyk K, Matas J. Forward-backward error: automatic detection of tracking failures. Proc 20th Int'l Conf Pattern Recogn 2010:23–6.
[21] Bouguet JY. Pyramidal implementation of the Lucas Kanade Feature Tracker description of the algorithm Technical Report. Intel Micro-processor Research Labs; 1999.
[22] Kalal Z, Matas J, Mikolajczyk K. P-N learning: bootstrapping binary classifiers by structural constraints. Conference on Computer Vision and Pattern Recognition, 2010.
[23] Kalal Z, Matas J, Mikolajczyk K. Online learning of robust object detectors during unstable tracking. On-line Learning for Computer Vision Workshop; 2009.
[24] Paganelli C, Lee D, Greer PB, et al. Quantification of lung tumor rotation with automated landmark extraction using orthogonal cine MRI images. Phys Med Biol 2015;60:7165–78.
[25] Dhont J, Vandemeulebroucke J, Burghelea M, et al. The long- and short-term variability of breathing induced tumor motion in lung and liver over the course of a radiotherapy treatment. Radiother Oncol 2018;126:339–46.
[26] Cusumano D, Dhont J, Boldrini L, et al. Predicting tumour motion during the whole radiotherapy treatment: a systematic approach for thoracic and abdominal lesions based on real time MR. Radiother Oncol 2018;129:456–62.
[27] Hu Y, Rankine L, Green OL, et al. Characterization of the onboard imaging unit for the first clinical magnetic resonance image guided radiation therapy system. Med Phys 2015;42:5828–37.
[28] Lowe DG. Distinctive image features from scale-invariant keypoints. Int J Comput Vis 2004;60:91–110.
[29] Cerviño LI, Du J, Jiang SB. MRI-guided tumor tracking in lung cancer radiotherapy. Phys Med Biol 2011;56:3773–85.
[30] Meyer P, Noblet V, Mazzara C, et al. Survey on deep learning for radiotherapy. Comput Biol Med 2018;98:126–46.
[31] Li P, Wang D, Wang L, Lu H. Deep visual tracking: review and experimental comparison. Pattern Recogn 2018;76:323–38.
[32] Wang L, Ouyang W, Wang X, Lu H. Visual tracking with fully convolutional networks. Proc IEEE ICCV 2015:3119–27.
[33] Paganelli C, Lee D, Kipritidis J, et al. Feasibility study on 3D image reconstruction from 2D orthogonal cine-MRI for MRI-guided radiotherapy. J Med Imaging Radiat Oncol 2018;62:389–400.
[34] Bjerre T, Crijns S, af Rosenschöld PM, et al. Three-dimensional MRI-linac intra-fraction guidance using multiple orthogonal cine-MRI planes. Phys Med Biol 2013;58:4943–50.