Evaluation of 2D and 3D ultrasound tracking algorithms and impact on ultrasound-guided liver radiotherapy margins

Valeria de Luca, Jyotirmoy Banerjee, Andre Hallack, Satoshi Kondo, Maxim Makhinya, Daniel Nouri, Lucas Royer, Amalia Cifor, Guillaume Dardenne, Orcun Goksel, et al.

To cite this version:

Valeria de Luca, Jyotirmoy Banerjee, Andre Hallack, Satoshi Kondo, Maxim Makhinya, et al.. Evaluation of 2D and 3D ultrasound tracking algorithms and impact on ultrasound-guided liver radiotherapy margins. Medical Physics, American Association of Physicists in Medicine, 2018, 45 (11), pp.4986-5003. 10.1002/mp.13152 . hal-01901201

HAL Id: hal-01901201

https://hal.archives-ouvertes.fr/hal-01901201

Submitted on 25 Oct 2018

Evaluation of 2D and 3D ultrasound tracking algorithms and impact on ultrasound-guided liver radiotherapy margins

Valeria De Luca
Computer Vision Laboratory, ETH Zurich, Zürich, Switzerland
Novartis Institutes for Biomedical Research, Basel, Switzerland

Jyotirmoy Banerjee
Translational Imaging Group, University College London, London, UK

Andre Hallack
Institute of Biomedical Engineering, University of Oxford, Oxford, UK

Satoshi Kondo
Konica Minolta Inc., Osaka, Japan

Maxim Makhinya
Computer Vision Laboratory, ETH Zurich, Zürich, Switzerland

Daniel Nouri
Natural Vision UG, Berlin, Germany

Lucas Royer
Institut de Recherche Technologique b-com, Rennes, France

Amalia Cifor
Mirada Medical, Oxford, UK

Guillaume Dardenne
Institut de Recherche Technologique b-com, Rennes, France

Orcun Goksel
Computer Vision Laboratory, ETH Zurich, Zürich, Switzerland

Mark J. Gooding
Mirada Medical, Oxford, UK

Camiel Klink
Department of Radiology, Erasmus MC, Rotterdam, The Netherlands

Alexandre Krupa, Anthony Le Bras and Maud Marchal
Institut de Recherche Technologique b-com, Rennes, France

Adriaan Moelker and Wiro J. Niessen
Department of Radiology, Erasmus MC, Rotterdam, The Netherlands

Bartlomiej W. Papiez
Institute of Biomedical Engineering, University of Oxford, Oxford, UK

Alex Rothberg
4Catalyzer Inc., Guilford, CT, USA

Julia Schnabel
School of Biomedical Engineering and Imaging Sciences, King's College London, London, UK

Theo van Walsum
Department of Radiology, Erasmus MC, Rotterdam, The Netherlands

Emma Harris
Institute of Cancer Research, London, UK

Muyinatu A. Lediju Bell
Department of Electrical and Computer Engineering, Johns Hopkins University, Baltimore, USA

Christine Tanner
Computer Vision Laboratory, ETH Zurich, Zürich, Switzerland

Purpose: Compensation for respiratory motion is important during abdominal cancer treatments. In this work we report the results of the 2015 MICCAI Challenge on Liver Ultrasound Tracking and extend the 2D results to relate them to clinical relevance in form of reducing treatment margins and hence sparing healthy tissues, while maintaining full duty cycle.
Methods: We describe methodologies for estimating and temporally predicting respiratory liver motion from continuous ultrasound imaging, used during ultrasound-guided radiation therapy. Furthermore, we investigated the trade-off between tracking accuracy and runtime in combination with temporal prediction strategies and their impact on treatment margins.
Results: Based on 2D ultrasound sequences from 39 volunteers, a mean tracking accuracy of 0.9 mm was achieved when combining the results from the 4 challenge submissions (1.2 to 3.3 mm). The two submissions for the 3D sequences from 14 volunteers provided mean accuracies of 1.7 and 1.8 mm. In combination with temporal prediction, using the faster (41 vs 228 ms) but less accurate (1.4 vs 0.9 mm) tracking method resulted in substantially reduced treatment margins (70% vs 39%) in contrast to mid-ventilation margins, as it avoided non-linear temporal prediction by keeping the treatment system latency low (150 vs 400 ms). Acceleration of the best tracking method would improve the margin reduction to 75%.
Conclusions: Liver motion estimation and prediction during free-breathing from 2D ultrasound images can substantially reduce the in-plane motion uncertainty and hence treatment margins. Employing an accurate tracking method while avoiding non-linear temporal prediction would be favorable. This approach has the potential to shorten treatment time compared to breath-hold and gated approaches, and increase treatment efficiency and safety.

Key words: image guidance, motion prediction, respiratory motion, treatment margins, ultrasound

1. INTRODUCTION

Intra-fraction organ motion due to breathing represents a challenge during intensity-modulated radiation therapy (IMRT) of the liver, lungs, pancreas, kidneys, breast and prostate.[1–8] The aim of IMRT is to deliver conformal and localized dose to the tumor, while sparing surrounding healthy tissue. Yet the motion of these organs requires substantially larger therapy margins (e.g. approximately 1–18 mm for lung, 10–55 mm for liver, 10–40 mm for kidney and 20–40 mm for pancreas[3]), to include the entire tumor volume in the treated area for the anticipated range of motion and hence, to ensure the effectiveness of the treatment.[3,9] Yet large margins are undesirable and reduce the advantages of IMRT.[10]

Image-guided radiation therapy uses imaging of target tissues prior to each fraction and may also provide continuous imaging during radiation delivery. This enables the estimation of the target position, size and shape, and the intra-fraction target motion. While most current treatment protocols attempt to arrest motion during radiation delivery using breath-hold, continuous imaging (or motion monitoring) can enable tracking of radiation beam to follow tumor motion in real-time.[11] One drawback of breath-holding is that repeat breath holds are often required if patients cannot hold their breath for the entire delivery period (i.e. greater than approximately 15 s[12]), thus treatment times will be lengthened as the beam is switched off between breath-holds. Also, it has been observed that the liver position is subject to variation between breath-holds and that the liver may undergo drift during breath-hold.[13–15] These uncertainties should be accounted for with an increase in treatment margins. Another option is gating the radiation beam whilst the patient breathes freely, which relies on a method of respiratory monitoring to determine the respiratory phase but target coverage can be compromised due to loss of correlation between internal motion and respiratory signals and irregular breathing and drift.[16] The advantage of continuous motion monitoring and tracking of the radiation beam over these techniques is that the treatment is not interrupted, patients can breath freely and that breathing irregularities and drift are compensated for dynamically rather than by increasing margins, which may result in a decrease in radiation-induced liver toxicity and duration of the therapy, and increase in chance of tumor control.[17,18]

Another frequently used technique for motion monitoring is the invasive implantation of fiducial markers for tumor tracking during radiation.[19–22] Fiducial markers can be used for image-based tracking, e.g. simultaneous kilovoltage and megavoltage imaging[23,24] or kilovoltage intra-fraction monitoring,[20] or non-image based tracking, e.g. as for radio frequency triangulation.[21,25] Examples of limitations of using fiducial markers are the requirement for surgery, possibility of markers migration and accuracy of the triangulation.[21,26]

Ultrasound (US) imaging is a suitable choice for observing motion during therapy due to its high temporal resolution, non-invasiveness and cost-efficiency. Currently, US-guided IMRT is mainly used in clinics to treat prostate[27,28] and breast cancer.[29] US-guided targeting of the liver in RT has been recently investigated.[30–32] However, during therapy fractions, liver tumors are not necessarily visible in US images. The acoustic impedance or acoustic reflectivity of liver tumors is often similar to that of surrounding tissue. This makes the tumors appear in US images with the same or similar echogenicity to tissue. The images can also be filled with acoustic clutter.[33] In both cases, it is difficult to distinguish the tumors in traditional US images.[34] Instead, the motion of other visible anatomical structures (e.g. vessels) can be estimated[35,36] and used as input to 4D liver motion models to spatially predict the tumor position.[37–40]

Although US probes are typically operated by hands, either passive arms or robotic arms can be used to hold the probe and therefore operators are not required to be in the treatment room. A robotic arm offers the additional advantage of moving with the target organ or helping inexperienced users to find it through cooperative control.[41,42]

Linear accelerator-based systems (LINAC) or adaptive targeting of the radiation beam using a multileaf collimator (MLC) or robotic treatment head, to follow the tumor motion during fractions, should take into consideration the treatment system latencies, including delays from the image acquisition, motion estimation algorithm, communication and control system, and beam delivery.[10,32,43] Therefore, the motion of the tumor should be accurately predicted for a sufficient time in the future to ensure the delivery of the radiation dose to the tumor and reduce the treatment margins.[44]

In this paper we investigate the impact of tracking liver motion under free breathing using US on treatment margins. US tracking has been investigated in several applications, e.g. for respiratory[35,36] and cardiac motion estimation.[36] However, reported performances in the liver are still not always suitable for direct translation into clinical application. In the case of respiratory motion, limiting factors are low robustness (i.e. high percentage of tracking errors >5 mm[3]) and high run-time (e.g. >300 ms) of the proposed algorithms, which both undermine the potential use of US tracking for online target localization during treatment. In addition, to the best of our knowledge, very few works have demonstrated its clinical impact in radiotherapy[5] and recent works on US guidance for real-time motion compensation are still based on phantom experiments.[45,46] Based on the results of the MICCAI 2015 Challenge on Liver Ultrasound Tracking (CLUST 2015) (http://clust.ethz.ch/), we propose an accurate and robust strategy to track anatomical landmarks in the liver, which fuses the tracking results of the algorithms that were submitted to the challenge. In addition to CLUST 2015, we temporally extrapolate the motion of the tracked landmarks, such as vessels, to compensate for system delays that occur in a real treatment scenario and investigate different strategies. Finally, the resulting uncertainties of the predicted motion are used to define motion-compensated treatment margins. These are compared to standard margins to investigate the efficiency of the proposed approach.

Compared to our previous benchmark (CLUST 2014),[35] we evaluated landmark tracking results on a larger dataset (overall +60% sequences and from 36 to 60 subjects) and with respect to more manual annotations, provided by three observers on 10% of the images. These annotations underwent a quality check and a correction if necessary to ensure optimal evaluation conditions. Furthermore, we provided a previously unseen validation set during the challenge, which was used to assess the tracking performance under realistic conditions and to evaluate the change of treatment margins when considering motion prediction. The aforementioned additions to the tracking challenge presented in this paper, i.e. temporal motion prediction and combined results for motion-compensated margin calculations, are novel compared to our previous study.[35]

The paper is organized as follows. In Section 2.A we describe the challenge data. In Section 2.B we list the tracking methods proposed by the challenge participant groups and their fusion. Temporal prediction was applied to selected tracking results, as described in Section 2.C. Predicted motion estimates were then used to compute treatment margins for motion-compensated therapy, see Section 2.D. Evaluation criteria are described in Section 2.E. Tracking and prediction results, and their impact in estimating treatment margins are reported and discussed in Sections III and IV. Finally, Section V summarizes conclusions of the challenge outcome and extension to motion-compensated treatment planning.

2. MATERIALS AND METHODS

2.A. Ultrasound data

A total of 85 US sequences of the liver of 60 healthy volunteers of 18 yr of age and older under free-breathing were collected between 2009 and 2015. Exclusion criteria were pregnancy, existing malignant tumor or undergoing cancer treatment. The data were provided by seven groups, the Biomedical Imaging Research Laboratory of CREATIS INSA, Lyon, France (CIL); Computer Vision Laboratory, ETH Zurich, Switzerland (ETH);[38,47] mediri GmbH, Heidelberg, Germany (MED); Biomedical Imaging Group, Departments of Radiology and Medical Informatics, Erasmus MC, Rotterdam, The Netherlands (EMC);[48] Joint Department of Physics, Institute of Cancer Research & Royal Marsden NHS Foundation Trust, London and Sutton, UK (ICR);[49,50] and SINTEF Medical Technology, Image Guided Therapy, Trondheim, Norway (SMT).[51] The acquisition and use of subject data were approved where applicable, by an ethics committee or institutional review board, and informed consent by each study participant was received.

An overview of the data is given in Appendix A. The sequences were acquired with a broad range of equipment (7 US scanners, 8 types of transducer) and different acquisition settings. The data consisted of 63 2D and 22 3D sequences from 42 and 18 subjects, respectively, which are characterized by duration ranging from 4 s to 10 min and temporal resolution form 6 to 31 Hz. Examples of the first frames and annotations are shown in Figure 1. Data were anonymized and randomly divided into two sets:

Training set: (40% of the sequences, i.e. 24 2D and seven 3D sequences), for which annotations of 10% of the images were released, to allow for tuning of the tracking algorithms.

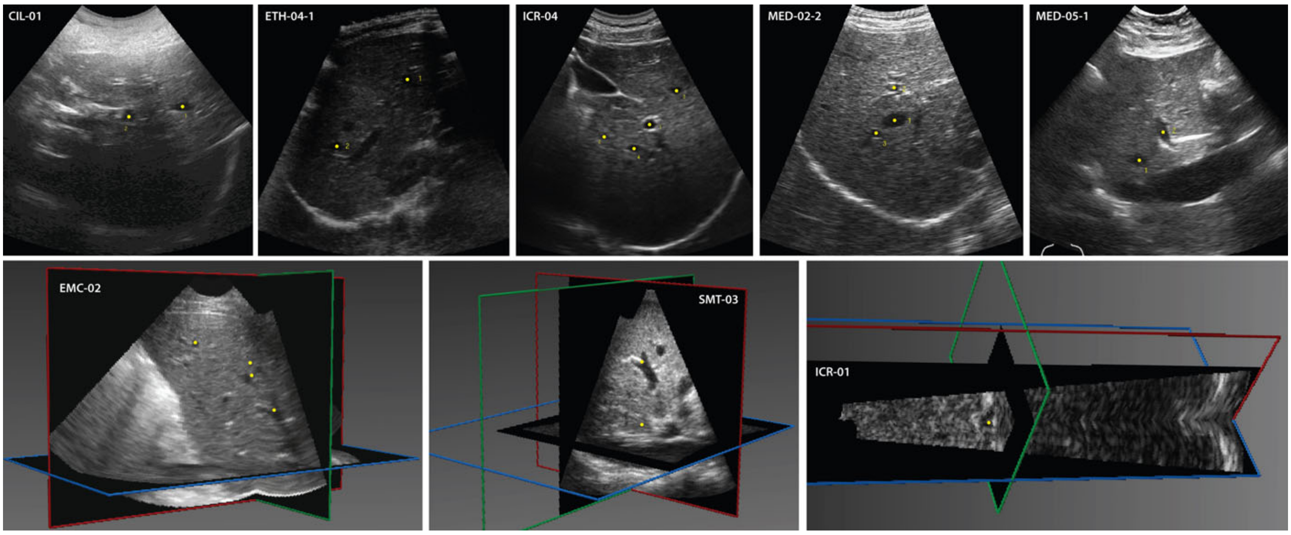

FIG. 1. Examples of first frame $I(1)$ of representative sequences of the training data: (top row) 2D sequences and (bottom row) 3D sequences. Point-landmarks $P_j(1)$ are highlighted and depicted in yellow.

Test set: (60%, 39 2D and 14 3D sequences), for which annotations of the first images were provided. The resulting tracking estimates were used to train the parameters of the temporal prediction model and to define treatment margins. From this set, 15 2D and 6 3D sequences were released during the on-site challenge at the MICCAI 2015 CLUST event.

2.B. Tracking methods

2.B.1. Tracking objective

Similarly to[35], $I(t,\mathbf{x})$ represents the intensity (or brightness) of the US image $I(t)$ at position $\mathbf{x}$, with $\mathbf{x} = [x_1, \ldots, x_D]^T \in \mathbb{R}^D$ for $D = \{2;3\}$, at frame t, with $t = 1,\ldots,T$ and T the total number of frames of the image sequence. The tracking objective was to compute the position of J point-landmarks $P_j(t) \in \mathbb{R}^D$ in each image, with $j = 1,\ldots,J$ and $J \in \{1;\ldots;5\}$ in this challenge, similar to.[35] The test set accounted a total $J_{TOT} =$ 85 2D and 22 3D landmarks. For all sequences, annotations of the first frame $P_j(1)$ was provided.

In the following we give an overview of the four 2D and two 3D tracking algorithms, which were submitted to the challenge. A detailed description of each method can be found in the challenge proceedings.[*]

2.B.2. 2D tracking

Nouri & Rothberg — *convolutional neural network:* The method proposed by *Nouri & Rothberg* trains a convolutional neural network to learn a function (with almost 1.9 millions parameters) which maps the intensities of image patches (46 × 46 pixels) into a low-dimensional embedding space, such that the Euclidean distance metric in that space is robust to the encountered landmark transformations.[52,53] A classical window search of size 24 pixels around a seed point p_j is used to find the location $P_j(t)$ in the new frame t that is most likely to be the tracked landmark. The window search finds the point that minimizes the learned distance metric to a template, which is composed of both the initial frame and the previous 10 frames $[I(1), I(t-10+1), \ldots, I(t)]$.

Kondo — *kernelized correlation filter: Kondo* proposed two extensions to the Kernelized correlation filter (KCF),[54] which applies a Gaussian kernel ($\sigma = 0.2$) to the correlation of image windows in the Fourier domain. KCF is extended by refining the initial tracked position by template matching in the region of ±2 pixels around the KCF prediction, based on normalized cross correlation. The second extension is the adaption of the window size. In a so-called calibration step, tracking is performed with a manually predetermined window size of 96 × 96 pixels for the first breathing cycle. Then the window size is revised based on the maximum frame-to-frame displacements and the feature size. From then onwards, the revised window size is used for KCF tracking with template matching refinement.

Makhinya & Goksel — *optical flow: Makhinya & Goksel* have extended an algorithm for identifying and tracking superficial veins in the forearm[55,56] by integrating several tracking recovery strategies to take advantage of the repetitive nature of respiration. Lucas-Kanade-based[57] tracking was applied on regularly-spaced grid points around each landmark, and used for *reference tracking* ($I(1)$ to $I(t)$), when the

[*] The proceedings of MICCAI 2015 CLUST are available at http://clust.ethz.ch/clust2015.html

local appearance of $I(1)$ and $I(t)$ is similar. Meanwhile, *iterative tracking* ($I(t-1)$ to $I(t)$) tracks points when the former fails. Each tracking strategy yields several motion vectors, which are then filtered for outliers. Subsequently, an affine transformation is fitted to the remaining vectors to provide a robust motion estimate for the landmark. For vessel-like structures *model-based tracking* is utilized via an axis-aligned ellipse representation of vessels. For each $I(t)$, first the ellipse is translated by the previous motion estimate. Then, its center and radii are re-estimated as in[56] using the Star edge detection, dynamic programming, model fitting, and binary templates. The resulting ellipse center is taken as the sought landmark.

Hallack, et al. — *logDemons:* Hallack et al. used the diffeomorphic logDemons image registration method,[58] but with a dense scale invariant feature transform (SIFT)[59] as similarity measure.[60, 61] Each landmark P_j was tracked independently and sequentially using image registration around a region of interest (W_j).[62] For frame $I(t)$, $W_{j,t-1}(t)$ of size 51×51 pixels is extracted around the previous estimated landmark location $P_j(t-1)$. The current landmark position $P_j(t)$ is obtained by finding a nonlinear transformation T_t between the moving ($W_{j,t-1}(t)$) and the fixed ($W_{j,1}(1)$) region. Both image regions were transformed into vector-valued images where a SIFT feature vector was computed for each pixel $\mathbf{x}$ from the histogram of gradients around its neighborhood (cell size of 2, 8 bins). The logDemons framework was applied using the sum of squared differences of the features vectors as image forces, 3 resolution levels, 6–20 iterations at each level and transformation field smoothing $\sigma_{\text{diff}} = 2$ pixels.

2.B.3. 3D tracking

Royer, et al. — *affine registration and mechanical model:* Royer et al. proposed a 3D target tracking method, which combines an intensity-based approach and mechanical regularization. The first step consists in obtaining a tetrahedral mesh model from a manual segmentation of the target in the initial image $I(1)$. To update the node positions of the model $\mathbf{q}(1)$ over time t, a dense displacement field is computed by minimizing the sum of squared intensity differences, for a piece-wise affine transformation model, using a steepest gradient optimization (step size $\alpha = 2 \times 10^{-6}$). To ensure robustness, each node displacement is constrained by internal forces of a mass-spring-damper system (spring stiffness 3.0, spring damping 1.0, nodal velocity damping 2.7), which is associated with the tetrahedral mesh.

Banerjee, et al. — *block matching and local registration:* The anatomical landmark tracking approach of Banerjee et al. consists of two rigid registration steps. In the global 4D tracking step, the whole liver volume is tracked by combining registrations to the previous and reference frame using the register-to-reference-by-tracking strategy.[63] This is followed by the local 3D registration step, where the tracking result from the previous step is refined by performing registration on the neighborhood region close to the anatomical landmark j, using the register-to-reference strategy.[64] Both steps use block-matching, with normalized cross correlation as similarity metric, followed by an outlier rejection scheme. Finally the rigid transformation is estimated from the trusted block-matching results.[64]

2.B.4. Decision fusion

To improve accuracy and robustness,[35,65] we combined the tracking results of all previously described methods (four 2D or two 3D methods) by computing for each frame t the median position of the tracked points $P_j(t)$ from these methods. Taking the median minimizes the absolute difference between the individual results and their combined value[66] and hence is robust to outliers (i.e. minority of predictions being bad). It assumes all results to be equally likely and should be extended to a weighted median if methods provide reliable uncertainties.

2.C. Temporal prediction for latency compensation

Conformal radiotherapy and IMRT treatments are usually delivered using MLCs. In image-guided dynamic MLC tracking, the accuracy of the target localization and treatment can be affected by the latency between the target motion and the MLC response.[67] The overall system latency is given by the sum of image formation, image processing and system adjustment latency. Image formation times depend on the US system and acquisition protocol, e.g. imaging depth and aperture, frame rate, dimensionality and transducer frequency. The latency due to image processing include run-times of motion estimation and temporal prediction algorithms. The duration for repositioning and adjusting a MLC ranges from 50 to 200 ms, depending on the target shift.[3,67–70] For example, approximately 50 ms are necessary to reposition the MLC for target shifts of 0.2–1.3 mm, 80 ms for 2 mm, and 200 ms for 5–6 mm.[67]

To compensate for the aforementioned system latency Δt, we forecast at time t the landmark position $P_j(t+\Delta t)$ at horizon Δt by using the available tracking positions $[P_j(1),\ldots,P_j(t)]$ and the temporal prediction approach proposed in Ref. [71] This approach considered the median results of four methods, namely a linear adaptive filter,[10] second order polynomial adaptive filter, support vector regression[72] and kernel density estimation,[73] as this previously provided improved results in comparison to the individual results.[71] Simulated annealing was used to optimize the method parameters for the motion trace of each P_j, based on leave-one-subject-out cross validation on the sequences in the test set. Finally, the method used the median parameters from the population excluding the subject. For short latencies ($\Delta t \leq 150$ ms) the linear adaptive filter provided similar results as the median (0.68 vs 0.62 mm) while having a much reduced run-time (1.3 vs 21.2 ms).[71] Hence both median fusion and linear adaptive filter were investigated in this work.

2.D. Treatment margins

Safety margins are required in the planning of all RT treatments, to compensate for known sources of error and ensure that the planned dose is delivered to the target.[74] Safety margins are added to the clinical target volume (CTV) and result in the planning target volume (PTV).[75]

Motion in the context of radiation therapy is defined as the displacement of target tissues between the planning CT image and treatment. Inter-fraction motion is the daily motion at the start of each treatment fraction, which can be minimized by correct patient positioning. Intra-fraction motion can occur during treatment due to patient movement or physiological processes, i.e. respiration. Errors in radiotherapy can be classified into systematic and random errors. For an individual patient, their systematic and random errors are their mean and standard deviation daily target displacements from their target position at planning (preparation) over all fractions, respectively. The population systematic and random errors are the standard deviation and root mean square of the individual patient systematic and random errors, respectively. Systematic errors occur in treatment preparation and include errors from set up (mainly due to patient positioning) and organ motion. Random errors occur during treatment and are caused by set up errors and organ motion.[74] The latter is comprised of inter-fraction motion, i.e. day-to-day motion, and intra-fraction motion, i.e. motion during treatment due for example to respiration.

Approximations have been proposed to estimate the margin sizes using so-called margin recipes, e.g.[76–78]. These generally assume that the magnitude of the motion is < 10 mm, errors <3 mm[77] and error components are Gaussian distributed and base the width of the margin on the sum of the variances of the contributory errors.[74] In this study we investigate the size of treatment margins required because of intra-fraction errors from respiratory motion. Inter-fraction errors, for example due to differences in patient positioning, are not investigated since the collected data does not provide this information.

2.E. Evaluation

We compared the performance of the tracking methods described in Section 2.B on the test set (see Appendix A), consisting of a total of 85 point-landmarks (e.g. vessel centers) in 39 2D sequences, and 22 point-landmarks (e.g. vessel bifurcations) in 14 3D sequences, which the observers were confident to be able to reliably annotate. In the following we describe the evaluation scheme used to validate and quantify the tracking and the prediction accuracy.

2.E.1 Tracking error

Three observers were asked to manually annotate the corresponding position of the initial point $P_j(1)$ in 10% of randomly selected images $I(\hat{t})$ from each sequence. The number of annotated frames/volumes per sequence is listed in Table A1. After review and eventual subsequent adjustment of the annotations by an additional observer, we computed the mean of the three annotations, denoted as $\hat{P}_j(\hat{t})$. Following the same error metrics as in,[35] the tracking error (TE) is calculated for each annotated frame $I(\hat{t})$ and landmark j as

$$TE_j(\hat{t}) = \|P_j(\hat{t}) - \hat{P}_j(\hat{t})\|, \tag{1}$$

where $\|.\|$ is the Euclidean distance between the estimated landmark position $P_j(\hat{t})$ and its mean manual annotation $\hat{P}_j(\hat{t})$. Results were then summarized by mean, standard deviation (Std) and 95th percentile of the single distribution including all $TE_j(\hat{t})$ belonging to a particular subgroup. These subgroups were the individual landmarks j, and landmark dimensionality (2D or 3D).

For baseline comparison and to estimate the motion magnitude of the landmarks, we included the case of no tracking, defined as

$$NoTE_j(\hat{t}) = \|P_j(1) - \hat{P}_j(\hat{t})\|. \tag{2}$$

2.E.2 Directional error

To assess the error in the main motion directions independently of the US probe orientation, we first determine the motion directions via principle component analysis (PCA) of each landmark trajectory, see Appendix B for details. The directional error is then computed as

$$DTE_{j,i}(\hat{t}) = p_{j,i}(\hat{t}) - \hat{p}_{j,i}(\hat{t}) \quad \in \mathbb{R}^D, \tag{3}$$

where $p_{j,i}(\hat{t})$ and $\hat{p}_{j,i}(\hat{t})$ are the projections of $\bar{P}_j(\hat{t})$ and $\hat{P}_j(\hat{t})$, respectively, onto the PCA space. Finally, we summarize the results by the mean and Std of the single distribution including all $DTE_{j,i}(\hat{t})$ belonging to a particular subgroup, as above. Directional errors (1st and 2nd PCA components) are reported only for 2D results, as these are used in this work to predict elliptical shaped treatment margins.

2.E.3 Prediction error

The error measures described in Sections 2.E.1 and 2.E.2 were also used to evaluate the prediction errors at time $\hat{t}^*$:

$$PE_j(\hat{t}^*) = \|PP_j(\hat{t}^*) - \hat{P}_j(\hat{t}^*)\| \tag{4}$$

and

$$DPE_{j,i}(\hat{t}^*) = pp_{j,i}(\hat{t}^*) - \hat{p}_{j,i}(\hat{t}^*), \tag{5}$$

where $PP_j(t^*) \in \mathbb{R}^D$ is the predicted position of landmark j at time $t^* = t + \Delta t$ (see Section 2.C) and $pp_{j,i}(t^*) \in \mathbb{R}$ is its projection in the i^{th} eigendirection.

The prediction performance was compared to the case of doing no temporal prediction, i.e. assuming no motion during Δt:

$$NoPE_j(\hat{t}^*) = \|P_j(t) - \hat{P}_j(\hat{t}^*)\|, \tag{6}$$

with $P_j(t)$ being the results of the considered tracking method.

2.E.4 Margin calculation

To illustrate the effect of the different methods on therapy margins, we employed a common recipe for calculating margins to compensate for intra-fraction motion errors, i.e. assuming zero set-up and delineation errors. Specifically, we used the population-based 3D margin recipe from van Herk et al.[77] given by:

$$\mathbf{m}_{PTV} = 2.5\mathbf{\Sigma} + 1.64(\boldsymbol{\sigma} - \boldsymbol{\sigma}_p) \approx 2.5\mathbf{\Sigma} + 0.7\boldsymbol{\sigma}', \tag{7}$$

where $\mathbf{\Sigma} = \sqrt{\mathbf{\Sigma}_m^2 + \mathbf{\Sigma}_s^2 + \mathbf{\Sigma}_d^2}$ denotes the Std of the systematic error, which is composed of the motion error Std ($\mathbf{\Sigma}_m$), the setup error Std ($\mathbf{\Sigma}_s$), and the delineation error ($\mathbf{\Sigma}_d$). The Std of the random error is given by $\boldsymbol{\sigma} = \sqrt{\boldsymbol{\sigma}_m^2 + \boldsymbol{\sigma}_s^2 + \boldsymbol{\sigma}_p^2}$, with motion error Std ($\boldsymbol{\sigma}_m$), setup error Std ($\boldsymbol{\sigma}_s$), and penumbra width Std ($\boldsymbol{\sigma}_p$). The approximation using $\boldsymbol{\sigma}' = \sqrt{\boldsymbol{\sigma}_m^2 + \boldsymbol{\sigma}_s^2}$ is valid for $\boldsymbol{\sigma}_p =$ 3.2 mm, $\boldsymbol{\sigma} \in [0,5]$ mm and big (diameter > 20 mm) CTVs of circular shape.[77] This recipe ensures that the CTV is fully covered by 95% of the prescribed dose for 90% of the patient population. As mentioned above, setup and delineation errors are unknown and hence set to zero in this work, i.e. $\mathbf{\Sigma}_s = \mathbf{\Sigma}_d = \boldsymbol{\sigma}_s = 0$.

The intra-fraction motion errors $\mathbf{\Sigma}_m$ and $\boldsymbol{\sigma}_m$ are determined from the mean and Std of the ith directional errors (DE) of J landmarks and K time points[76] via

$$M_i = \frac{1}{J}\sum_{j=1}^{J}\mu_{j,i}, \qquad \Sigma_{m,i} = \sqrt{\frac{1}{J}\sum_{j=1}^{J}(\mu_{j,i} - M_i)^2} \tag{8}$$

and

$$\sigma_{m,i} = \sqrt{\frac{1}{J}\sum_{j=1}^{J}\sigma_{j,i}^2}, \tag{9}$$

where

$$\mu_{j,i} = \sum_{k}^{K} DE_{j,i}(\hat{t}_k^*)/K, \tag{10}$$

$$\sigma_{j,i} = \sqrt{\sum_{k}^{K}(DE_{j,i}(\hat{t}_k^*) - \mu_{j,i})^2/K} \tag{11}$$

and DE stands for either DTE or DPE as defined in Eqs. (3) and (5).

3. RESULTS

3.A. 2D tracking

The results of the 2D point-landmark tracking on the test set are summarized in Table I. The mean TE ranges from 1.2 mm to 3.4 mm for the methods submitted to MICCAI 2015 CLUST, with best results achieved by Hallack et al. Fusing the results of all tracking methods improved accuracy by 24–73% in comparison to the individual results. Yet these errors are higher than the inter-observer variability, with mean (95%) TE of the three observers <0.5 (1.1) mm. The tracking error distributions are shown in Figure 2.

To assess the robustness of all methods, we quantified the percentage of failures on the test set, i.e. the percentage of landmark results for which TE > 3 mm or TE > 5 mm, see Table I. The fusion method achieved the highest robustness, with TE > 5 mm in only 1.0% of the landmark results and TE > 3 mm in 4.3%.

There was low correlation between the motion magnitude of the landmarks and the tracking errors, with the sample Pearson correlation coefficients ρ ranging from 0.01 to 0.25. Correlation between TE and imaging center frequency (see Table A1, surrogate measure for image quality) was found to be low for all methods ($\rho \in [0.06,0.18]$). We also found low correlation between the tracking error (TE) and Std of the observers' error ($\rho \in [0.02,0.12]$).

The mean run-time per frame, determined per sequence, was per method at most 41 ms (*Makhinya & Goksel*) to 228 ms (*Kondo*), see Table I. The tracking method of *Makhinya & Goksel* was faster than the US acquisition frame-rate for all sequences.

3.B. 3D tracking

The results of 3D tracking on the test set are shown in Table II. On average, the highest accuracy and fastest runtime were achieved by Royer et al., with TE of 1.7 ± 0.9 mm and 350 ms per volume. The percentages of failures, for which TE > 3 mm, ranged between 8.4% and 8.7%, while for only 0.8–1.7% of annotated landmarks TE was <5 mm.

As before, we investigated if the tracking error is correlated with the motion of the landmarks and found weak sample Pearson correlation coefficients $\rho = 0.29$ and 0.39 for Royer et al. and Banerjee et al., respectively. Low correlation was also found between TE and Std of the observers TE for Royer et al. ($\rho = 0.36$) and Banerjee et al. ($\rho = 0.11$). Moderate correlation ($\rho = 0.44$ and 0.48) was only found between tracking errors and center frequency of the imaging acquisition (see Table AI) for each landmark j.

3.C. Temporal prediction

Forecasting of the landmark motion traces was evaluated only for 2D sequences, as these have a sufficient amount of long sequences to adapt the temporal prediction model. Furthermore, run-times of the proposed 2D tracking methods are short enough to not create a great burden for the prediction,

TABLE I. Results of **2D** tracking on the test set. Best results among the automatic methods are highlighted in **bold**. (D)TE: (directional) tracking error; 1st (2nd): error in the first (second) direction of motion.

	TE (mm)				Landmarks (%)		1st DTE (mm)		2nd DTE (mm)		
Method	Mean	Std	95%	Mean TE range (mm)	TE > 3 mm	>5 mm	Mean	Std	Mean	Std	Run-time (ms)
No tracking	6.45	5.11	16.48	[2.76, 17.06]	69.72	51.09	−0.87	7.99	0.35	1.72	–
Nuori & Rothberg	3.35	5.21	14.19	[0.46, 23.03]	27.36	18.58	1.44	5.14	0.49	3.10	100
Kondo	2.91	10.52	5.18	[**0.33**, 56.21]	9.32	5.19	−1.32	10.20	−0.49	3.63	228
Makhinya & Goksel	1.44	2.80	3.62	[0.49, 16.67]	5.60	3.75	0.28	2.75	0.32	1.47	**41**
Hallack et al.	1.21	3.17	2.82	[0.34, 16.13]	4.63	2.18	**0.09**	3.29	**0.04**	0.85	208
Fusion	**0.92**	**0.98**	**2.78**	[0.34, **3.52**]	**4.31**	**0.95**	0.23	**1.15**	0.13	**0.65**	228
Observer 1	0.46	0.36	1.13	[0.22, 1.28]	0.09	0.00	−0.06	0.44	−0.08	0.37	–
Observer 3	0.47	0.34	1.08	[0.23, 1.33]	0.07	0.00	0.05	0.44	−0.01	0.38	–
Observer 2	0.44	0.32	1.03	[0.21, 1.20]	0.04	0.00	0.01	0.42	0.09	0.33	–

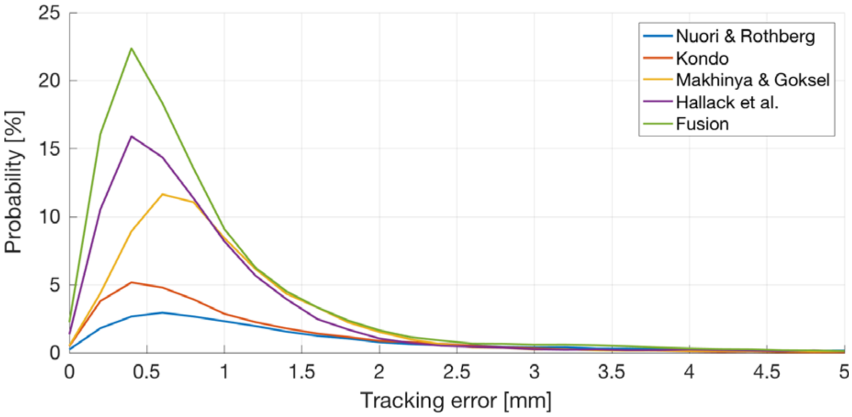

FIG. 2. 2D tracking error distributions up to 5 mm.

TABLE II. Results of **3D** tracking on the test set. Best results among the automatic methods are highlighted in **bold**.

	Tracking error (mm)				Landmarks (%) with		
Method	Mean	Std	95%	Mean TE range (mm)	TE > 3 mm	TE > 5 mm	Run-time (ms)
No tracking	5.54	3.77	12.17	[1.95, 10.66]	68.35	48.10	–
Royer et al.	**1.74**	**0.92**	3.65	[0.79, **3.54**]	8.86	0.84	**350**
Banerjee et al.	1.80	1.64	**3.41**	[0.78, 5.24]	8.44	1.69	10860
Fusion	1.74	1.15	3.49	[**0.78**, 3.71]	8.44	0.84	10860
Observer 3	1.36	1.14	3.37	[0.55, 4.68]	6.33	2.53	–
Observer 1	1.27	1.07	3.47	[0.33, 3.25]	9.28	0.84	–
Observer 2	1.19	0.83	2.89	[0.44, 2.40]	4.64	0.00	–

and hence allow envisioning clinical applicability of such a guidance system. On the contrary, the fastest 3D tracking approach (Royer et al.) requires 350 ms and 3D acquisition (with a large enough field of view to reliably capture respiratory motion) and MLC tracking add another approximately 170 ms[68] to the system latency. Hence the prediction horizon needs to be greater than 500 ms. In addition, the short 3D sequences, which were available for this study, do not allow to train all temporal prediction methods. Therefore we did not include 3D forecasting in this work. 3D guidance from 2D motion predictions can be achieved by carefully aligning the US transducer with the main direction of the respiratory motion, calibrating the US coordinate system with the treatment coordinate system, and employing a 4D motion model.[38,40]

Figure 3 compares the mean prediction error (MPE) of the linear adaptive filter and decision fusion prediction. Predictions are obtained from leave-one-subject-out cross-validation on

Fusion (the most accurate) and *Makhinya & Goksel* (the fastest) tracking results on 2D test data. When considering each tracking strategy, using temporal prediction resulted in lower errors than without prediction for all horizons $\Delta t \in \{150,300,400,600,1000\}$ ms, apart from the result of Fusion for $\Delta t = 150$ ms, where MPE = 1.44 (1.43) mm for median (no) prediction. Linear adaptive filters achieved the highest accuracy for short latencies ($\Delta t = 150$ ms, MPE = 1.10 mm for Fusion tracking). For $\Delta t = 300$ ms results are very similar, while for higher Δt the median-based prediction outperforms the linear filter by 9% to 44% (3% to 49%) for Fusion (*Makhinya & Goksel*) tracking. In all cases, errors increased with latency, suggesting that shorter latencies are preferable.

3.D. Treatment margins and strategies

We investigate the trade-off between accuracy and run-time by comparing the treatment margins required for the following three motion compensation strategies: **A.** fusion tracking and median prediction method (total run-time 251 ms) with 400 ms latency; **B.** fast tracking by *Makhinya & Goksel* and linear adaptive filter (run-time 42 ms) with 150 ms latency; and **C.** fusion tracking and linear filter (run-time 235 ms) with 150 ms latency. Even though fusion tracking would require a speed-up by a factor of 5.6 for strategy **C** to become feasible, we included it to investigate the potential benefit of such a speed-up. The latencies chosen for strategies **A** and **B** include algorithm run-time, and re-positioning and adjustments time of the treatment beam. We considered approximately 150 ms and 100 ms for strategies **A** and **B**, respectively, as times that are needed to adjust the treatment beam.[67] These latencies increase with the magnitude of the target position shift (see Section 2.D), which depends on the time difference between consecutive position estimations, given by the tracking algorithm run-time. The three strategies are compared to (a) only tracking with the two aforementioned approaches (no prediction); (b) the mid-ventilation approach,[79,80] where the time-weighted mean position of the tracked landmark $\bar{P}_j$ over the first three breathing cycles is used as landmark location throughout therapy; and (c) the case of no motion compensation, i.e. without tracking. Results are summarized in Table III.

Table III also lists the safety margins $\mathbf{m}_{PTV} = \{m_{i,PTV}\}$ due to respiration, computed as described in Eqs. (7)and (8) in the main motion directions $i \in \{1;2\}$. Compared to the baseline (no tracking), $\mathbf{m}_{PTV}$ can be reduced by 62–84% and 29–69% in each direction when using tracking without and with temporal prediction, respectively. For the mid-ventilation approach, $\mathbf{m}_{PTV}$ reduction is 14–32% (Fusion tracking) and 13–31% (*Makhinya & Goksel* tracking method) in each direction. Strategies characterized by lower error variance, such as **C**, result in smaller margins. Figure 4 illustrates the moving margin ellipse for one representative landmark and breathing cycle, and compares Strategy **C** to the baseline of no tracking and mid-ventilation. It can be observed that the fixed margins (no tracking, mid-ventilation) require larger margins to not miss the moving vessel center, while Strategy **C** is able to stay close to it. The population-based margins do not ensure that all targets are fully encompassed by the PTV 100% of the time, such as in end-inhale positions (see $t = 23.70$ s in Figure 4, bottom row).

In addition, we considered a spherical CTV with 50 mm diameter, representing the central cross-section of a stage T2 to T3a liver tumor,[81] and added the resulting elliptical-shape 2D margins. The PTV is then the ellipsoid with semi-axes $[m_{1,\ PTV},\ m_{2,\ PTV},\ 0]$ + CTV radius. The margin volume ($V_{\mathbf{m}} = PTV - CTV$) is reduced from 73,390 mm^3 (no tracking) to 51,294 mm^3 (margin volume reduction $rV_{\mathbf{m}} = 30\%$) when employing the mid-ventilation margins (based on the mid-position from Fusion tracking), and to 13,234 mm^3 (82%) and 12,649 mm^3 (83%) by strategy **C** without and with temporal prediction, respectively.

To incorporate motion prediction in the out-of-plane direction, we considered the same 3rd component of the ellipsoidal margin of size $m_{3,PTV} = 3$ mm[39] for all strategies, see last column of Table III. The margin volume is

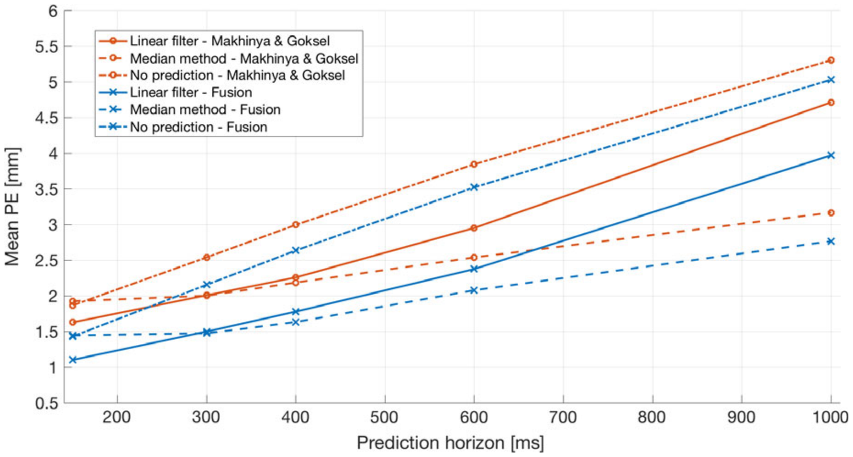

FIG. 3. Summary of the mean prediction error vs prediction horizon (Δt) for the test set in case of tracking based on Fusion or *Makhinya & Goksel* combined with no prediction, linear or median prediction.

TABLE III. Summary of prediction errors (PE), directional prediction errors (DPE) in the first (1st) and second (2nd) direction of motion, 2D margins ($\mathbf{m}_{PTV}$) and margin volume reduction relative to no tracking for 2D margins ($rV_{\mathbf{m}}$) and when considering an additional fixed margin in the 3rd direction $m_{3,PTV}$ = 3 mm ($rV_{\mathbf{m},3D}$). Three selected motion-compensation strategies are compared, namely: **A**. slower, more accurate tracking and 400 ms latency; **B**. faster, less accurate tracking and 150 ms latency; **C**. same accurate tracking as strategy **A**, but accelerated 5.3 times and 150 ms latency (currently unfeasible) on the test set. The mid-ventilation results for strategy **C** are the same as for **A**, as both use the same tracking method.

	PE (mm)				1st DPE (mm)		2nd DPE (mm)		$\mathbf{m}_{PTV}$ (mm)			
Strategy	Mean	Std	95%	Mean PE range (mm)	Mean	Std	Mean	Std	1st D	2nd D	$rV_{\mathbf{m}}$ (%)	$rV_{\mathbf{m},3D}$ (%)
No tracking	6.13	4.59	15.20	[2.76, 13.79]	**–0.16**	14.65	0.54	3.06	19.20	5.00	0	0
A												
No prediction	2.63	1.97	6.20	[1.19, 6.31]	0.23	3.18	0.13	0.80	4.52	1.66	−77	−70
Prediction	1.63	1.46	4.28	[0.60, 4.57]	0.26	2.01	0.15	0.82	3.76	1.74	−79	−73
Mid-ventilation	5.13	4.09	13.35	[1.98, 14.06]	0.31	6.34	0.49	1.56	13.05	4.30	−30	−27
B												
No prediction	1.87	2.84	4.52	[0.79, 17.09]	0.27	3.02	0.32	1.50	7.17	3.35	−59	−54
Prediction	1.63	2.89	4.23	[0.64, 18.43]	0.28	2.91	0.31	1.54	7.39	3.50	−57	−52
Mid-ventilation	5.27	4.13	13.61	[2.16, 13.55]	0.34	6.46	0.72	1.57	13.32	4.33	−29	−26
C												
No prediction	1.43	1.20	3.57	[0.65, 4.01]	0.23	1.71	0.13	**0.70**	3.30	**1.55**	−82	−75
Prediction	**1.10**	**1.08**	**3.09**	[**0.46**, **3.70**]	0.25	**1.34**	**0.12**	0.72	**3.03**	1.61	**–83**	**–76**

now reduced from 90,072 mm^3 (no tracking) to 65,308 mm^3 (3D margin volume reduction $rV_{\mathbf{m},3D}$ = 27%) when using the mid-position from Fusion tracking; 24,748 mm^3 (73%) when using strategy **A** with temporal prediction; and 22,031 mm^3 (76%) when using strategy **C** with temporal prediction.

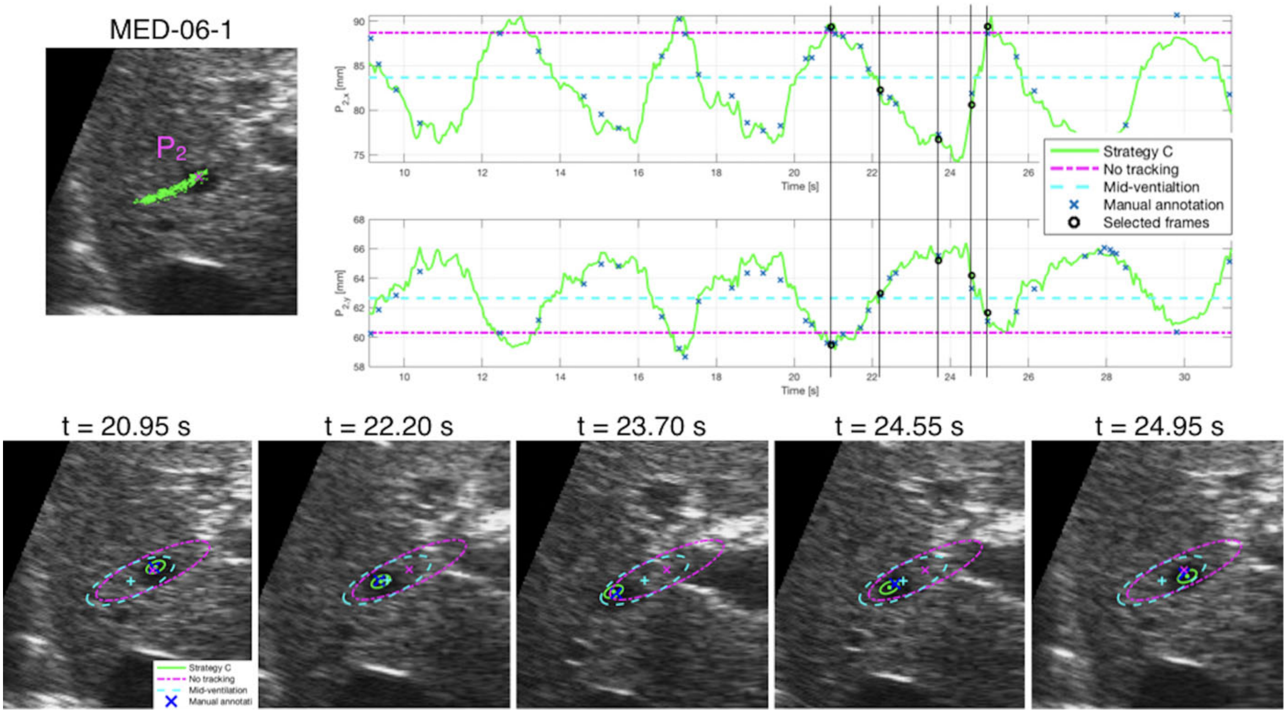

FIG. 4. Illustration of prediction results with margins for a representative case (MPEj = 1:44 mm). (Top row, left) fixed region from reference image $I(1)$ with magenta cross for $\hat{P}_j(1)$ and green dots showing the predicted positions $P_j(t)$ by strategy **C** for 5 consecutive breathing cycles. (Top row, right) plot of corresponding results from strategy **C** (green line) vs initial (dash-dot magenta line) and mid-ventilation positions (dashed cyan line) over time, with black circles highlighting 5 results shown in details. Manually annotated $\hat{P}_j(\hat{t})$ are displayed as blue crosses. (Bottom row) Images for the 5 highlighted results. Green ellipse represents margin required for strategy **C**, centered in predicted position $P_j(\hat{t})$ (green dot). Magenta dash-dotted ellipse shows margin required for no tracking strategy and hence is centered at $\hat{P}_j(1)$ (magenta cross). The fixed margin of the mid-ventilation strategy is depicted as dashed line and plus symbol in cyan.

4. DISCUSSION

4.A. Respiratory motion estimation

When comparing individual 2D tracking methods submitted to the 2015 MICCAI CLUST competition (see Sections 2.B), the combination of optical-flow and model based tracking proposed by *Makhinya & Goksel* was the only real-time approach, while ranking second in terms of accuracy. Its main strength is the combination of several fast tracking strategies. The highest accuracy was achieved by the diffeomorphic log-Demons image registration by *Hallack et al.*, by using dense features and having the most flexible transformation model. The KCF in the Fourier space of *Kondo* had the highest run-time and failed in tracking one sequence. The CNN-based block-matching method from *Nuori & Rothberg* had the overall lowest accuracy and robustness, with a mean TE of 3.4 mm and 19% of failures. While learning a suitable similarity measure by the CNN has potentials, its drawback is the need for a large dataset of manual annotations for training.

Computing the median value of the results from all 2D tracking methods yielded on average the highest landmark tracking accuracy (0.9 mm) and robustness (<1% failures). This reduced mean motion by 86%, but required the run-time of at least the slowest method (228 ms). When comparing to the previously published 2D tracking results of CLUST 2014,[35] mean errors of the individual submitted methods were in a similar range, i.e. from 1.4 to 2.1 mm. Four out of 6 compared individual approaches could achieve real-time performance. Similar to this work, considering the median tracking results of all submitted methods improved the mean tracking error to 1.2 mm and 1.6% failures, which are higher than the presented results. Yet this comparison is only qualitative, as the CLUST 2014 evaluation was based on a smaller dataset and slightly different manual annotations. Recently, after the on-site MICCAI 2015 CLUST event, Shepard et al. validated a GPU implementation of a learning-based block-matching approach on the CLUST 2015 data.[82] Compared to the proposed median fusion approach, this method achieved real-time performance (4–14 vs 228 ms), lower mean tracking error (0.72 vs 0.92 mm) but higher error Std (1.25 vs 0.98 mm). Future work should extend results to the latest submitted methods and evaluate their impact on margin reductions.

For spatially predicting the 3D position of the treatment target, our 2D motion estimation and temporal prediction results could be used as input to a 4D liver model.[37,39] Such an approach achieved *in-vivo* for the right liver lobe of eight subjects a spatio-temporal 3D mean prediction accuracy of 2.4 (2.7) mm for a system latency of 150 (400) ms based on 2D US tracking (mean accuracy 0.9 mm) and a population 4D motion model.[38] Note that improved performance was observed for the central liver region, which is closer to the typical US plane, and by adapting the model to the subject using few breath-hold observations.[40]

For 3D tracking, the combination of an intensity-based affine registration and a mechanical model proposed by Royer et al. was computationally much more efficient, but still far from being real-time, than the block matching algorithm of Banerjee et al. The latter had a run time of almost 11 s, which limits its applicability in clinical practice. Accuracy of these two methods was comparable (MTE = 1.7 and 1.8 mm) and greater than two of the 2D algorithms. Furthermore, averaging the tracking results did not improve performance. Fusing results from more methods might improve accuracy. These results are generally improved compared to the ones of CLUST 2014, where a 3D mean tracking error between 2.5 and 4.6 mm on a smaller dataset was reported.[35]

The tracking performance was generally not dependent on the motion magnitude of 2D landmarks nor image quality. When considering the 3D case, the moderate correlations of tracking error with motion magnitude and image quality, together with lower volume rates, suggest that advances in 4D image resolution could improve results. Another aspect to take into account is the difficulty of visually inspecting and annotating 3D sequences. This is supported by the higher intra-observer variability of the manual annotations in 3D, as shown by the mean TE of the observers in Table II (mean $TE_{3\mathrm{D}} \in [1.19, 1.36]$ mm) vs Table I (mean $TE_{2\mathrm{D}} \in [0.44, 0.47]$ mm). This difference is still substantial even when approximately adjusting for differences due to 2D vs 3D measures, i.e. mean $TE^{3\mathrm{D}}_{2\mathrm{D}} \in [0.53, 0.58]$ mm, where $TE^{3\mathrm{D}}_{2\mathrm{D}}(\hat{t}) = \sqrt{3\, TE^{2}_{2\mathrm{D}}(\hat{t})/2}$ which assumes equal error components.[35]

A limitation of this work is that we were not able to quantify potential errors that may be introduced by the US guidance system. Ultrasound-based tracking of hepatic vessels for the purpose of radiotherapy relies on accurate spatial calibration, which enables transformation of pixel locations in the ultrasound image to treatment room coordinates. Geometrical accuracy and precision of US localization of <1 mm can be achieved,[83] however care must be taken to ensure the calibration is subject to strict quality assurance.[84] Another source of error is speed of sound (SOS) error.[85] Most US systems assume a fixed SOS of 1540 ms^{-1} for all tissues, however SOS varies with tissue type potentially causing inaccurate measurement of the change in depth of a hepatic vessel as the subject breathes. Assuming a SOS in liver of 1595 ms^{-1}[86] and a maximum axial vessel displacement of 20 mm, SOS will give an maximum error in vessel motion of approximately 0.7 mm. Methods for the correction of SOS error have been explored in depth by Fontanarosa et al.[85] and future work should incorporate such approaches.

In a real treatment scenario, delays of the treatment system introduce errors in the estimation of the target position during therapy. Hence we quantified the prediction errors using two approaches proposed in,[71] namely a fast linear adaptive filter and a median fusion of four linear and non-linear methods, see Section 2.C. These approaches were tested on the fastest (*Makhinya & Goksel*) and most accurate (Fusion) tracking results. For the four compared combinations, prediction errors increased by 39% to 73% with prediction horizons from 150 to 1000 ms, see Figure 3. Despite being currently unfeasible, the lowest error (mean PE = 1.1 mm) is obtained by strategy **C** (Fusion tracking + linear adaptive filter for

150 ms latency). When comparing strategy **B** (based on a fast tracking method, which requires shorter prediction horizons) to **A** (more accurate tracking with longer horizon), the mean prediction errors are similar while the Std almost doubled for strategy **B**, leading to substantially increased margins, see Section 4.B. This investigation shows that the trade-off between a faster, less accurate versus a slower, more accurate tracking method can be hard to judge and should be quantified in the system context.

The tracking and prediction performances were worse than the observer annotation accuracy, indicating the potential for additional improvements. The American Association of Physicists in Medicine recommends for external-beam radiation therapy the use of respiratory management when the target motion is greater than 5 mm.[3] Therefore, errors of target prediction should also not exceed 5 mm. Overall this was not yet achieved by any automatic method and observers had difficulties in 3D. Yet for the most accurate 2D and 3D tracking method, TE was greater than 5 mm in only 0.95% and 0.84% of the 2D and 3D images, respectively. With lower tracking errors and much lower number of failures, results have substantially improved compared to those of CLUST 2014.[35] For 2D prediction, errors were greater than 5 mm in 3.25%, 4.19%, and 1.31% of the images for strategies **A**, **B** and **C**, respectively.

4.B. Impact on treatment margins

We investigated how the trade-off between accuracy and speed of tracking methods affects the treatment system performance, by combining tracking with temporal prediction to compensate for the corresponding system latency. The system performance was quantified by means of required treatment margins to compensate for intra-fraction motion. This showed that for a tumor of 50 mm diameter, the most accurate tracking method combined with temporal prediction (strategy **A**) reduces the volume of healthy tissue which gets irradiated by 79% in comparison to no tracking. Trading tracking accuracy against speed (strategy **B**) was counterproductive, with margin volumes being only reduced by 57%. Speeding-up the accurate tracking to achieve strategy **C** provided another reduction by 4% over strategy **A**, which amounts to 2425 mm^3 spared healthy tissue. Margin reductions are much lower for the mid-ventilation approach (30%) than for any tracking method.

Margin size generally increases with DPE. An exception is strategy **B**, no prediction versus strategy **A**, no prediction, which have similar error in the first motion component of the prediction errors (1st DPEs) (0.27 ± 3.02 vs 0.23 ± 3.18mm) but quite different margins (7.17 vs 4.52 mm). This comes from large differences in the standard deviations of the systematic error $\Sigma_{m,1}$ (2.11 vs 0.89 mm) between the two methods. The discrepancies in ranges of mean PE ([0.79, 17.09] vs [1.19, 6.31] mm) and Std PE (2.84 vs 1.97 mm) hinted at such an effect. Hence tracking and prediction errors should be assessed not only with regard to their mean errors but also to their error variation.

Respiratory motion per image sequence followed a Gaussian distribution only in 24.7%. In fact, respiratory motion follows per cycle a path similar to a sin[48] function, which is not normally distributed.[87] In addition, landmark positions can drift over time, influencing the motion distribution and leading to multi-modal distributions. After tracking, residual errors were more often Gaussian distributed (62.3%) and 95th percentiles substantially reduced. Based on results from,[87,88] which showed the applicability of the margin recipe for similar distributions and motion amplitudes below 10 mm, we conclude that the margin recipe can be applied to the tracked data.

Our analysis has focused on 2D tissue tracking, which, considering the limited validation options and increased image acquisition and processing times of 3D tracking, may be the more practical to implement. Liver motion is typically greatest in the superior-inferior (SI) and anterior-posterior (AP) directions with left-right (LR) motion being significantly smaller. In some studies, typical liver motion in the LR direction is reported to be <2 mm[89,90] and therefore relatively small PTV margins to account for this motion could be applied in the LR directions. This relies on the accurate alignment of the 2D transducer with the SI/AP plane, which may be assisted by optical tracking of the transducer. To incorporate motion uncertainties in the out-of-plane direction, we considered the same margin size $m_{3,PTV} = 3$ mm for all strategies in Table III. Slightly smaller margin reductions were obtained: from 76% to 73% for strategy **A** with temporal prediction, from 83% to 76% for strategy **C** with temporal prediction, and from 26–27% to 29–30% for mid-ventilation for added $m_{3,PTV} = 3$ mm. Note that for these margin calculations zero out-of-plane motion was assumed ($m_{3,PTV} = 0$). Increasing $m_{3,PTV}$ leads to larger margins and reduced margin reductions (73% to 76%, mid-ventilation 26–27% to 29–30% for $m_{3,PTV} = 3$ mm). However the ranking of the strategies remains the same[†]. As in other studies,[87,91] we neglect setup errors as we focus on intra-fraction motion mitigation. Margins of all strategies will need to be increased by the same amount to compensate for setup errors. Similarly as for out-of-plane motion, this will increase the required margins, but ranking of the strategies would remain the same.

5. CONCLUSIONS

In this work, we compared different tracking and prediction techniques, and combined results in a common margin

[†] The margin volume reduction of method a over method b is given by

$$rV_{\mathbf{m}}^{(a,b)} = \frac{V_{PTV}^{(a)} - V_{CTV}}{V_{PTV}^{(b)} - V_{CTV}} - 1 = \frac{s r_{1,PTV}^{(a)} r_{2,PTV}^{(a)} - V_{CTV}}{s r_{1,PTV}^{(b)} r_{2,PTV}^{(b)} - V_{CTV}} - 1,$$

where $r_{i,PTV} = m_{i,PTV} + r_{i,CTV}$, $s = 4\pi/3\ r_{3,PTV}$, and the same $m_{3,PTV}$ is added in the 3rd dimension. The ranking of the methods remain the same for all $s>0$, as $rV_{\mathbf{m}}^{(a,b)} < rV_{\mathbf{m}}^{(c,b)}$ if $r_{1,PTV}^{(a)}\ r_{2,PTV}^{(a)} < r_{1,PTV}^{(c)}\ r_{2,PTV}^{(c)}$.

recipe to predict treatment margins for ultrasound-guided radiation therapy of the liver.

We first validated tracking on a large dataset of 2D and 3D US sequences of the liver of volunteers under free breathing. We compared several approaches, as part of the 2015 MICCAI CLUST workshop. The tracking of anatomical landmarks achieved an overall accuracy of 0.9 and 1.7 mm for 2D and 3D sequences, respectively. In 2D, the best results were obtained by fusing all proposed algorithms by computing the median estimation per frame. In 3D the combination of intensity-based registration and mechanical model of Royer et al. obtained the highest accuracy.

Adding temporal prediction is fundamental to compensate for the treatment system latency and hence to correctly compensate for the target motion during therapy. Therefore, we compared two 2D prediction approaches, namely using a fast linear adaptive filter or the median of four linear and non-linear methods. These were applied to the 2D tracking results. Both high tracking accuracy and short prediction horizon (i.e. high computational speed) positively influence the accuracy of temporal prediction. The lowest prediction error of 1.1 mm was achieved by strategy **C**. Given the high run-time of the fusion algorithm, this strategy is not yet feasible. Yet computational improvements and implementation optimization could reduce the current run-time.

Accurate compensation for target motion results in a potential reduction of 79% to 83% of the treatment margin volume. The mid-ventilation strategy could reduce this treatment margin volume by only 29–30%.

The proposed tracking and prediction approach can be applied to US guidance in IMRTs to continuously estimate the motion of the organ under treatment and hence reduce treatment margins, which decreases dose to healthy tissue. Due to a duty cycle of 100% compared to gating with the patient free-breathing or in breath-holding, this approach would allow for shorter and more efficient treatments.

Future work includes testing of the motion compensation framework on patient data, use of 4D motion models[39] of the liver for the spatial (and hence spatio-temporal) prediction of the tumor motion.

ACKNOWLEDGMENTS

V De Luca and C Tanner acknowledge funding from the European Union's 7th Framework Program (FP7/2007-2013) under grant agreement No. 611889 (TRANS-FUSIMO). M A Lediju Bell acknowledges funding from the UNCF/Merck Postdoctoral Science Research Fellowship. A Hallack et al. acknowledges funding from the CRUK/EPSRC Cancer Imaging Centre at Oxford; the Research Council UK Digital Economy Programme under grant N^{o} EP/G036861/1 (Oxford Centre for Doctoral Training in Healthcare Innovation); and CAPES Foundation, process N^{o} BEX 0725/12-9. The challenge organizers express appreciation to Frank Lindseth and Sinara Vijayan from SINTEF Medical Technology, Trondheim, Norway; Adeline Bernard and Hervé Liebgott from CREATIS INSA, Lyon, France; Julia Schwaab from mediri GmbH, Heidelberg, Germany, for providing data and annotations. We also thank Alicia B. Dagle and Anastasia K. Ostrowski (Johns Hopkins University, Baltimore, USA), Tuathan O'Shea and Srikanta Sharma (Institute of Cancer Research, London, UK), Alba M. Malavé Dos Santos and Ina Kompan for annotating part of the data.

CONFLICT OF INTEREST

The authors have no conflicts to disclose.

APPENDIX A

DATA

Tables AI–AII list all relevant information on the 2D and 3D US sequences, respectively, namely: size and length, spatial and temporal resolutions, number of annotations and US scanner details. We used the following convention to assign sequence names: InstitutionAbbreviation-InstitutionSubjectNumber-RepetitionNumber. For example, ETH-01-2 corresponds to the US sequence of subject no. 1 and repetition number 2 provided by ETH Zurich. When the repetition number is not provided, only one sequence was acquired for the correspondent subject.

Table AI. Summary of the **2D** challenge data (part 1 of 2). The sequence name (first column) of the test set is listed in regular black font. The training sequences, for which all available annotations were provided, are highlighted in bold font. The test data provided at the on-site challenge are underlined.

	Sequence info				Annotation		Acquisition info		
Sequence	Im.size (pix/vox)	Im.res. (mm)	No. frames	Im.rate (Hz)	No. ann.	No.ann. frames	Scanner	Probe	Center freq. (MHz)
CIL-01	480 × 640	0.30	1342	22	2	144	Ultrasonix MDP	4DC7-3/40	4.5
CIL-02	480 × 640	0.40	1075	17	1	131	Ultrasonix MDP	4DC7-3/40	4.5
CIL-03	480 × 640	0.40	1070	18	2	138	Ultrasonix MDP	4DC7-3/40	4.5

Table AI. Continued.

	Sequence info				Annotation		Acquisition info		
Sequence	Im.size (pix/vox)	Im.res. (mm)	No. frames	Im.rate (Hz)	No. ann.	No.ann. frames	Scanner	Probe	Center freq. (MHz)
CIL-04	480 × 640	0.50	895	15	2	112	Ultrasonix MDP	4DC7-3/40	4.5
CIL-05	480 × 640	0.30	1430	23	2	161	Ultrasonix MDP	4DC7-3/40	4.5
ETH-01-1	490 × 570	0.40	3652	15	2	366	Siemens Antares	CH4-1	2.22
ETH-01-2	482 × 608	0.41	4650	15	2	466	Siemens Antares	CH4-1	2.22
ETH-02-1	472 × 565	0.42	2620	15	1	263	Siemens Antares	CH4-1	2.22
ETH-02-2	462 × 590	0.41	4826	15	1	483	Siemens Antares	CH4-1	2.22
ETH-03-1	473 × 437	0.28	4588	14	1	460	Siemens Antares	CH4-1	2.22
ETH-03-2	464 × 442	0.28	4191	13	1	420	Siemens Antares	CH4-1	2.22
ETH-04-1	469 × 523	0.40	5247	16	2	525	Siemens Antares	CH4-1	1.82
ETH-04-2	480 × 652	0.38	4510	14	2	452	Siemens Antares	CH4-1	1.82
ETH-05-1	462 × 563	0.42	4615	15	2	463	Siemens Antares	CH4-1	1.82
ETH-05-2	477 × 556	0.40	3829	13	2	384	Siemens Antares	CH4-1	1.82
ETH-06-1	462 × 580	0.40	5244	16	1	525	Siemens Antares	CH4-1	2.00
ETH-06-2	476 × 604	0.38	5165	16	1	518	Siemens Antares	CH4-1	2.00
ETH-07-1	475 × 548	0.37	5586	17	2	560	Siemens Antares	CH4-1	1.82
ETH-07-2	467 × 568	0.37	5582	17	2	559	Siemens Antares	CH4-1	1.82
ETH-08-1	466 × 562	0.36	5574	17	2	558	Siemens Antares	CH4-1	1.82
ETH-08-2	466 × 589	0.36	5577	17	2	559	Siemens Antares	CH4-1	1.82
ETH-09-1	464 × 560	0.40	4587	15	4	460	Siemens Antares	CH4-1	1.82
ETH-09-2	479 × 566	0.42	4590	15	3	460	Siemens Antares	CH4-1	1.82
ETH-10-1	462 × 589	0.36	5578	17	3	559	Siemens Antares	CH4-1	1.82
ETH-10-2	470 × 595	0.36	5584	17	3	559	Siemens Antares	CH4-1	1.82
ETH-11-1	478 × 552	0.45	4284	14	2	429	Siemens Antares	CH4-1	2.22
ETH-11-2	476 × 541	0.45	3785	12.4	1	380	Siemens Antares	CH4-1	2.22
ETH-12-1	264 × 313	0.71	14516	25	1	1453	Siemens Antares	CH4-1	2.22
ETH-12-2	262 × 313	0.77	15640	25	1	1565	Siemens Antares	CH4-1	2.22
ETH-13-1	268 × 304	0.71	9934	25	1	994	Siemens Antares	CH4-1	2.00
ETH-13-2	268 × 304	0.71	10525	25	1	1054	Siemens Antares	CH4-1	2.00
ICR-01	393 × 457	0.55 × 0.42	4858	23	3	608	Elekta Clarity, Ultrasonix	m4DC7-3/40	4.5
ICR-02	393 × 457	0.55 × 0.42	3481	23	2	436	Elekta Clarity, Ultrasonix	m4DC7-3/40	4.5
ICR-03	393 × 457	0.55 × 0.42	3481	23	3	436	Elekta Clarity, Ultrasonix	m4DC7-3/40	4.5
ICR-04	393 × 457	0.55 × 0.42	3481	23	4	349	Elekta Clarity, Ultrasonix	m4DC7-3/40	4.5
ICR-05	397 × 485	0.55 × 0.43	3481	20	2	348	Elekta Clarity, Ultrasonix	m4DC7-3/40	4.5
ICR-06	397 × 485	0.55 × 0.43	3481	21	2	348	Elekta Clarity, Ultrasonix	m4DC7-3/40	4.5
ICR-07	397 × 495	0.49 × 0.38	3481	23	2	348	Elekta Clarity, Ultrasonix	m4DC7-3/40	4.5
ICR-08	399 × 495	0.50 × 0.39	3481	23	3	348	Elekta Clarity, Ultrasonix	m4DC7-3/40	4.5
ICR-09	399 × 485	0.57 × 0.44	3481	19.9	2	349	Elekta Clarity, Ultrasonix	m4DC7-3/40	4.5
ICR-10	397 × 495	0.49 × 0.38	3481	23.5	2	349	Elekta Clarity, Ultrasonix	m4DC7-3/40	4.5
MED-01-1	408 × 512	0.41	2455	20	3	246	DiPhAs Fraunhofer	VermonCLA	5.5
MED-02-1	408 × 512	0.41	2458	20	3	246	DiPhAs Fraunhofer	VermonCLA	5.5
MED-02-2	408 × 512	0.41	2443	20	3	245	DiPhAs Fraunhofer	VermonCLA	5.5
MED-02-3	408 × 512	0.41	2436	20	5	244	DiPhAs Fraunhofer	VermonCLA	5.5
MED-03-1	408 × 512	0.41	2442	20	2	245	DiPhAs Fraunhofer	VermonCLA	5.5
MED-03-2	408 × 512	0.41	2450	20	3	246	DiPhAs Fraunhofer	VermonCLA	5.5
MED-04-1	524 × 591	0.35	3304	11	1	331	Zonare z.one	C4-1	4.0
MED-05-1	524 × 591	0.35	3304	11	2	331	Zonare z.one	C4-1	4.0
MED-06-1	408 × 512	0.41	2427	20	4	243	DiPhAs Fraunhofer	VermonCLA	5.5
MED-06-2	408 × 512	0.41	2424	20	3	243	DiPhAs Fraunhofer	VermonCLA	5.5
MED-07-1	408 × 512	0.41	2470	20	3	248	DiPhAs Fraunhofer	VermonCLA	5.5
MED-07-2	408 × 512	0.41	2478	20	3	248	DiPhAs Fraunhofer	VermonCLA	5.5

TABLE AI. Continued.

	Sequence info				Annotation		Acquisition info		
Sequence	Im.size (pix/vox)	Im.res. (mm)	No. frames	Im.rate (Hz)	No. ann.	No.ann. frames	Scanner	Probe	Center freq. (MHz)
MED-07-3	408 × 512	0.41	2450	20	3	246	DiPhAs Fraunhofer	VermonCLA	5.5
MED-07-4	408 × 512	0.41	2456	20	4	246	DiPhAs Fraunhofer	VermonCLA	5.5
MED-08-1	524 × 591	0.35	3304	11	3	331	Zonare z.one	C4-1	4.0
MED-08-2	524 × 591	0.35	3304	11	3	331	Zonare z.one	C4-1	4.0
MED-09	408 × 512	0.48	2420	30	1	243	DiPhAs Fraunhofer	VermonCLA 3.5	3.4
MED-10	408 × 512	0.45	2416	31	2	243	DiPhAs Fraunhofer	VermonCLA 3.5	3.4
MED-11	408 × 512	0.45	2425	31	2	243	DiPhAs Fraunhofer	VermonCLA 3.5	3.4
MED-12	408 × 512	0.48	2415	30	2	242	DiPhAs Fraunhofer	VermonCLA 3.5	3.4
MED-13	475 × 687	0.27	3135	17	1	314	Zonare z.one	C6-2	~4.0
MED-14	475 × 687	0.27	3855	17	2	386	Zonare z.one	C6-2	~4.0

TABLE AII. Summary of the **3D** challenge data. The sequence name (first column) of the test set is listed in regular black font. The training sequences, for which all available annotations were provided, are highlighted in bold font. The test data provided at the on-site challenge are underlined.

	Sequence info				Annotation		Acquisition info		
Sequence	Im.size (pix/vox)	Im.res. (mm)	No. frames	Im.rate (Hz)	No. ann.	No.ann. frames	Scanner	Probe	Center freq. (MHz)
EMC-01	192 × 246 × 117	1.14 × 0.59 × 1.19	79	6	1	8	Philips iU22	X6-1	3.2
EMC-02	192 × 246 × 117	1.14 × 0.59 × 1.19	54	6	4	6	Philips iU22	X6-1	3.2
EMC-03	192 × 246 × 117	1.14 × 0.59 × 1.19	159	6	1	16	Philips iU22	X6-1	3.2
EMC-04	192 × 246 × 117	1.14 × 0.59 × 1.19	140	6	1	15	Philips iU22	X6-1	3.2
EMC-05	192 × 246 × 117	1.14 × 0.59 × 1.19	147	6	1	15	Philips iU22	X6-1	3.2
EMC-06-1	192 × 246 × 117	1.14 × 0.59 × 1.19	100	6	1	11	Philips iU22	X6-1	3.2
EMC-06-2	192 × 246 × 117	1.14 × 0.59 × 1.19	100	6	1	11	Philips iU22	X6-1	3.2
EMC-06-3	192 × 246 × 117	1.14 × 0.59 × 1.19	100	6	1	11	Philips iU22	X6-1	3.2
EMC-07-1	192 × 246 × 117	1.14 × 0.59 × 1.19	100	6	1	11	Philips iU22	X6-1	3.2
EMC-07-2	192 × 246 × 117	1.14 × 0.59 × 1.19	100	6	1	11	Philips iU22	X6-1	3.2
EMC-07-3	192 × 246 × 117	1.14 × 0.59 × 1.19	100	6	1	11	Philips iU22	X6-1	3.2
ICR-01	480 × 120 × 120	0.31 × 0.51 × 0.67	141	24	1	15	Siemens SC2000	4Z1c	2.8
ICR-02	480 × 120 × 120	0.31 × 0.51 × 0.67	141	24	1	20	Siemens SC2000	4Z1c	2.8
SMT-01	227 × 227 × 229	0.70	97	8	3	96	GE E9	4V-D	2.5
SMT-02	227 × 227 × 229	0.70	96	8	3	92-93	GE E9	4V-D	2.5
SMT-03	227 × 227 × 229	0.70	96	8	2	45-96	GE E9	4V-D	2.5
SMT-04	227 × 227 × 229	0.70	97	8	1	96	GE E9	4V-D	2.5
SMT-05	227 × 227 × 229	0.70	96	8	2	64-96	GE E9	4V-D	2.5
SMT-06	227 × 227 × 229	0.70	97	8	3	49-96	GE E9	4V-D	2.5
SMT-07	227 × 227 × 229	0.70	97	8	2	95	GE E9	4V-D	2.5
SMT-08	227 × 227 × 229	0.70	97	8	3	96	GE E9	4V-D	2.5
SMT-09	227 × 227 × 229	0.70	97	8	3	96	GE E9	4V-D	2.5

APPENDIX B

PCA

We used Principal Components Analysis (PCA) to compute the directional tracking error, see Section 2.E.2. For each tracking method and landmark j, PCA consists in solving the eigenproblem:

$$\mathbf{C}\mathbf{w}_i = \lambda_i \mathbf{w}_i, \qquad \forall i \in [1, \ldots, D], \tag{13}$$

where $\mathbf{C} = [\bar{\hat{P}}_j(1), \ldots, \bar{\hat{P}}_j(\hat{T})]^T \times [\bar{\hat{P}}_j(1), \ldots, \bar{\hat{P}}_j(\hat{T})]$ is the covariance matrix of the centered manual annotations of the D-dimensional landmark $j\bar{\hat{P}}_j(\hat{t}) = \hat{P}_j(\hat{t}) - \pi_j$, with $\pi_j = 1/\hat{T} \sum_{\hat{t}=1}^{\hat{T}} \hat{P}_j(t)$ the mean position and $\hat{T}$ the number of annotated frames for landmark j. λ_i are the sorted eigenvalues ($\lambda_i > \lambda_{i+1}$) and $\mathbf{w}_i$ the corresponding eigenvectors. For each i^{th} eigendirection, we calculated the trajectory projection $\hat{p}_{j,i}(\hat{t}) = \mathbf{w}_i^T \hat{\hat{P}}_j(\hat{t})$. Similarly, we then project the tracking trajectories P_j onto the PCA space by $p_{j,i}(\hat{t}) = \mathbf{w}_i^T (P_j(\hat{t}) - \pi_j)$.

V. De Luca, E. Harris, M.A. Lediju Bell and C. Tanner organized MICCAI CLUST 2015; all other authors contributed results of their tracking methods; Website: http://clust.ethz.ch/

[a)] Author to whom correspondence should be addressed: Electronic mail: valeria.de_luca@novartis.com.

REFERENCES

1. Bortfeld T, Jokivarsi K, Goitein M, Kung J, Jiang SB. Effects of intra-fraction motion on IMRT dose delivery: statistical analysis and simulation. *Phys Med Biol*. 2002;47:2203.
2. Jiang SB, Pope C, Al Jarrah KM, Kung JH, Bortfeld T, Chen GT. An experimental investigation on intra-fractional organ motion effects in lung IMRT treatments. *Phys Med Biol*. 2003;48:1773.
3. Keall PJ, Mageras GS, Balter JM, et al. The management of respiratory motion in radiation oncology report of AAPM Task Group 76a). *Med Phys*. 2006;33:3874–3900.
4. Omari EA, Erickson B, Ehlers C, et al. Preliminary results on the feasibility of using ultrasound to monitor intrafractional motion during radiation therapy for pancreatic cancer. *Med Phys*. 2016;43:5252–5260.
5. OShea T, Bamber J, Fontanarosa D, van der Meer S, Verhaegen F, Harris E. Review of ultrasound image guidance in external beam radiotherapy part II: intra-fraction motion management and novel applications. *Phys Med Biol*. 2016;61:R90.
6. Ozhasoglu C, Murphy MJ. Issues in respiratory motion compensation during external-beam radiotherapy. *Int J Radiat Oncol Biol Phys*. 2002;52:1389–1399.
7. Shirato H, Seppenwoolde Y, Kitamura K, Onimura R, Shimizu S. Intrafractional tumor motion: lung and liver. *Semin Radiat Oncol*. 2004;14:10–18.
8. Webb S. Motion effects in (intensity modulated) radiation therapy: a review. *Phys Med Biol*. 2006;51:R403.
9. Morgan-Fletcher SL. Prescribing, recording and reporting photon beam therapy (supplement to ICRU report 50). ICRU report 62, PP. IX+52. *Br J Radiol*. 1999;74:294.
10. Vedam S, Keall P, Docef A, Todor D, Kini V, Mohan R. Predicting respiratory motion for fourdimensional radiotherapy. *Med Phys*. 2004;31:2274–2283.
11. Verellen D, De Ridder M, Linthout N, Tournel K, Soete G, Storme G. Innovations in image-guided radiotherapy. *Nat Rev Cancer*. 2007;7:949–960.
12. Boda-Heggemann J, Knopf AC, Simeonova-Chergou A, et al. Deep inspiration breath holdbased radiation therapy: a clinical review. *Int J Radiat Oncol Biol Phys*. 2016;94:478–492.
13. Eccles C, Brock KK, Bissonnette JP, Hawkins M, Dawson LA. Reproducibility of liver position using active breathing coordinator for liver cancer radiotherapy. *Int J Radiat Oncol Biol Phys*. 2006;64:751–759.
14. Korreman SS. Motion in radiotherapy: photon therapy. *Phys Med Biol*. 2012;57:R161.
15. Parkes MJ, Green S, Cashmore J, et al. *Int J Radiat Oncol Biol Phys*. 2016;96:709–710.
16. Korreman SS, Juhler-N∅ttrup T, Boyer AL. Respiratory gated beam delivery cannot facilitate margin reduction, unless combined with respiratory correlated image guidance. *Radiother Oncol*. 2008;86:61–68.
17. Dawson LA, Jaffray DA. Advances in image-guided radiation therapy. *J Clin Oncol*. 2007;25:938–946.
18. Ehrbar S, Perrin R, Peroni M, et al. Respiratory motion-management in stereotactic body radiation therapy for lung cancer – a dosimetric comparison in an anthropomorphic lung phantom (luca). *Radiother Oncol*. 2016;121:328–334.
19. Depuydt T, Poels K, Verellen D, et al. Treating patients with real-time tumor tracking using the vero gimbaled linac system: implementation and first review. *Radiother Oncol*. 2014;112:343–351.
20. Iwata H, Ishikura S, Murai T, et al. A phase i/ii study on stereotactic body radiotherapy with real-time tumor tracking using CyberKnife based on the Monte Carlo algorithm for lung tumors. *Int J Clin Oncol*. 2017;22:706–714.
21. Shirato H, Harada T, Harabayashi T, 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 Rad Oncol Biol Phys*. 2003;56:240–247.
22. Takao S, Miyamoto N, Matsuura T, et al. Intrafractional baseline shift or drift of lung tumor motion during gated radiation therapy with a real-time tumor-tracking system. *Int J Radiat Oncol Biol Phys*. 2016;94:172–180.
23. Hunt MA, Sonnick M, Pham H, et al. Simultaneous MV–kV imaging for intrafractional motion management during volumetric-modulated arc therapy delivery. *J Appl Clin Med Phys*. 2016;17:473–486.
24. Wiersma R, Mao W, Xing L. Combined kV and MV imaging for real-time tracking of implanted fiducial markers. *Med Phys*. 2008;35:1191–1198.
25. Harris EJ, Donovan EM, Yarnold JR, Coles CE, Evans PM. Characterization of target volume changes during breast radiotherapy using implanted fiducial markers and portal imaging. *Int J Radiat Oncol Biol Phys*. 2009;73:958–966.
26. Kitamura K, Shirato H, Shimizu S, et al. Registration accuracy and possible migration of internal fiducial gold marker implanted in prostate and liver treated with real-time tumor-tracking radiation therapy (RTRT). *Radiother Oncol*. 2002;62:275–281.
27. Fung AY, Ayyangar KM, Djajaputra D, Nehru RM, Enke CA. Ultrasound-based guidance of intensity-modulated radiation therapy. *Med Dosim*. 2006;31:20–29.
28. Lattanzi J, McNeeley S, Hanlon A, Schultheiss TE, Hanks GE. Ultrasound-based stereotactic guidance of precision conformal external beam radiation therapy in clinically localized prostate cancer. *Urology*. 2000;55:73–78.
29. Fontanarosa D, van der Meer S, Bamber J, Harris E, OShea T, Verhaegen F. Review of ultrasound image guidance in external beam radiotherapy: I. treatment planning and inter-fraction motion management. *Phys Med Biol*. 2015;60:R77.
30. Fuss M, Salter BJ, Cavanaugh SX, et al. Daily ultrasound-based image-guided targeting for radiotherapy of upper abdominal malignancies. *Int J Radiat Oncol Biol Phys*. 2004;59:1245–1256.
31. Bloemen-van Gurp E, van der Meer S, Hendry J, et al. Active breathing control in combination with ultrasound imaging: a feasibility study of image guidance in stereotactic body radiation therapy of liver lesions. *Int J Radiat Oncol Biol Phys*. 2013;85:1096–1102.
32. OShea TP, Bamber JC, Harris EJ. Temporal regularization of ultrasound-based liver motion estimation for image-guided radiation therapy. *Med Phys*. 2016;43:455–464.
33. Lediju MA, Pihl MJ, Dahl JJ, Trahey GE. Quantitative assessment of the magnitude, impact and spatial extent of ultrasonic clutter. *Ultrason Imaging*. 2008;30:151–168.
34. Cho SH, Lee JY, Han JK, Choi BI. Acoustic radiation force impulse elastography for the evaluation of focal solid hepatic lesions: preliminary findings. *Ultrasound Med Biol*. 2010;36:202–208.
35. De Luca V, Benz T, Kondo S, König L, et al. The 2014 liver ultrasound tracking benchmark. *Phys Med Biol*. 2015;60:5571.
36. De Luca V, Székely G, Tanner C. Estimation of large-scale organ motion in B-mode ultrasound image sequences: a survey. *Ultrasound Med Biol*. 2015;41:3044–3062.
37. McClelland JR, Hawkes DJ, Schaeffter T, King AP. Respiratory motion models: a review. *Med Image Anal*. 2013;17:19–42.
38. Preiswerk F, De Luca V, Arnold P, et al. Model-guided respiratory organ motion prediction of the liver from 2D ultrasound. *Med Image Anal*. 2014;18:740–751.

39. Tanner C, Boye D, Samei G, Szekely G. Review on 4D models for organ motion compensation. *Crit Rev Biomed Eng*. 2012;40:135–154.
40. Tanner C, Yang M, Samei G, Székely G. Influence of inter-subject correspondences on liver motion predictions from population models. In: *Int Symposium on Biomedical Imaging*. Prague, Czech Republic: IEEE; 2016:286–289.
41. Şen, HT, Bell MAL, Zhang Y, et al. System integration and in vivo testing of a robot for ultrasound guidance and monitoring during radiotherapy. *IEEE Trans Biomed Eng*. 2017;64:1608–1618.
42. Su L, Iordachita I, Zhang Y, et al. Feasibility study of ultrasound imaging for stereotactic body radiation therapy with active breathing coordinator in pancreatic cancer. *J Appl Clin Med Phys*. 2017;18:84–96.
43. Sharp GC, Jiang SB, Shimizu S, Shirato H. Prediction of respiratory tumour motion for real-time imageguided radiotherapy. *Phys Med Biol*. 2004;49:425.
44. Verma PS, Wu H, Langer MP, Das IJ, Sandison G. Survey: real-time tumor motion prediction for imageguided radiation treatment. *Comp Sci Engin*. 2011;13:24–35.
45. Ipsen S, Bruder R, OBrien R, Keall PJ, Schweikard A, Poulsen PR. Online 4D ultrasound guidance for real-time motion compensation by MLC tracking. *Med Phys*. 2016;43:5695–5704.
46. Schwaab J, Prall M, Sarti C, et al. Ultrasound tracking for intra-fractional motion compensation in radiation therapy. *Phys Med*. 2014;30:578–582.
47. De Luca V, Tschannen M, Székely G, Tanner C. A learning-based approach for fast and robust vessel tracking in long ultrasound sequences. In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI*. Vol. 8149. Berlin, Heidelberg: Springer; 2013:518–525.
48. Banerjee J, Klink C, Peters ED, Niessen WJ, Moelker A, van Walsum T. 4D liver ultrasound registration. In: *Biomedical Image Registration*, Cham: Springer; 2014:194–202.
49. Bell MAL, Byram BC, Harris EJ, Evans PM, Bamber JC. In vivo liver tracking with a high volume rate 4D ultrasound scanner and a 2D matrix array probe. *Phys Med Biol*. 2012;57:1359.
50. Lediju M, Byram BC, Harris EJ, Evans PM, Bamber JC, et al. 3D liver tracking using a matrix array: implications for ultrasonic guidance of IMRT. In: *Ultrasonics Symposium*. San Diego, CA: IEEE; 2010:1628–1631.
51. Vijayan S, Klein S, Hofstad EF, Lindseth F, Ystgaard B, Lango T. Validation of a non-rigid registration method for motion compensation in 4D ultrasound of the liver. In: International Symposium on Biomedical Imaging. Melbourne, Australia: IEEE; 2013:792–795.
52. Chopra S, Hadsell R, LeCun Y. Learning a similarity metric discriminatively, with application to face verification. In: *Computer Vision and Pattern Recognition*. Vol. 1. San Diego, CA: IEEE; 2005:539–546.
53. Hadsell R, Chopra S, LeCun Y. Dimensionality reduction by learning an invariant mapping. In: *Computer Vision and Pattern Recognition*. Vol. 2. New York, NY: IEEE; 2006:1735–1742.
54. Henriques JF, Caseiro R, Martins P, Batista J. High-speed tracking with kernelized correlation filters. *Trans Pattern Anal Mach Intell*. 2015;37:583–596.
55. Crimi A, Makhinya M, Baumann U, Szekely G, Goksel O. Vessel tracking for ultrasound-based venous pressure measurement. In: *IEEE Int Symp Biomedical Imaging*. 2014:306–9.
56. Crimi A, Makhinya M, Baumann U, Thalhammer C, Szekely G, Goksel O. Automatic measurement of venous pressure using B-mode ultrasound. *IEEE Trans Biomedical Engineering*. 2016;63:288–299.
57. Lucas B, Kanade T. An iterative image registration technique with an application to stereo vision. In: *Proceedings of Imaging Understanding Workshop*; 1981:121–130.
58. Vercauteren T, Pennec X, Perchant A, Ayache N. Symmetric log-domain diffeomorphic registration: a demons-based approach. In: *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2008*. New York:Springer; 2008:754–761.
59. Liu C, Yuen J, Torralba A. SIFT flow: Dense correspondence across scenes and its applications. *Trans Pattern Anal Mach Intell*. 2011;33:978–994.
60. Cifor A, Risser L, Chung D, Anderson EM, Schnabel J, et al. Hybrid feature-based diffeomorphic registration for tumor tracking in 2-D liver ultrasound images. *IEEE Tran Med Imaging*. 2013;32:1647–1656.
61. Hallack A, Papież BW, Wilson J, et al. Correlating tumour histology and *ex vivo* MRI using dense modality-independent patch-based descriptors. In: *International Workshop on Patch-based Techniques in Medical Imaging*; 2015:129–136.
62. König L, Kipshagen T, Rühaak J. A non-linear image registration scheme for real-time liver ultrasound tracking using normalized gradient fields. *Challenge on Liver Ultrasound Tracking CLUST 2014*; 2014:29.
63. Banerjee J, Klink C, Niessen W, Moelker A, van Walsum T. 4D ultrasound tracking of liver and its verification for TIPS guidance. *IEEE Trans Med Imaging*. 2015;35:52–62.
64. Banerjee J, Klink C, Peters ED, Niessen WJ, Moelker A, van Walsum T. Fast and robust 3D ultra sound registration – block and game theoretic matching. *Med Image Anal*. 2015;20:173–183.
65. Sinha A, Chen H, Danu D, Kirubarajan T, Farooq M. Estimation and decision fusion: a survey. *Neurocomputing*. 2008;71:2650–2656.
66. Yager RR, Rybalov A. Understanding the median as a fusion operator. *Int J Gen Syst*. 1997;26:239–263
67. Poulsen PR, Cho B, Sawant A, Ruan D, Keall PJ. Detailed analysis of latencies in image-based dynamic MLC tracking. *Med Phys*. 2010;37:4998–5005.
68. Goodband J, Haas O, Mills J. A comparison of neural network approaches for on-line prediction in IGRT. *Med Phys*. 2008;35:1113–1122.
69. Jin JY, Yin FF. Time delay measurement for linac based treatment delivery in synchronized respiratory gating radiotherapy. *Med Phys*. 2005;32:1293–1296.
70. Ren Q, Nishioka S, Shirato H, Berbeco RI. Adaptive prediction of respiratory motion for motion compensation radiotherapy. *Phys Med Biol*. 2007;52:6651.
71. Tanner C, Eppenhof K, Gelderblom J, Szekely G. Decision fusion for temporal prediction of respiratory liver motion. In: *International Symposium on Biomedical Imaging*. Beijing, China: IEEE; 2014:698–701.
72. Riaz N, Shanker P, Wiersma R, Gudmundsson O, Mao W, Widrow B, Xing L. Predicting respiratory tumor motion with multi-dimensional adaptive filters and support vector regression. *Phys Med Biol*. 2009;54:5735.
73. Ruan D. Kernel density estimation-based real-time prediction for respiratory motion. *Phys Med Biol*. 2010;55:1311
74. Van Herk M. Errors and margins in radiotherapy. *Semin Radiat Oncol*. 2004;14:52–64.
75. van Herk M, Remeijer P, Lebesque JV. Inclusion of geometric uncertainties in treatment plan evaluation. *Int J Radiat Oncol Biol Phys*. 2002;52:1407–1422.
76. Cacicedo J, Perez J, de Zarate RO, et al. A prospective analysis of inter- and intrafractional errors to calculate CTV to PTV margins in head and neck patients. *Clin Transl Oncol*. 2015;17:113–120.
77. van Herk M, Remeijer P, Rasch C, Lebesque JV. The probability of correct target dosage: dose-population histograms for deriving treatment margins in radiotherapy. *Int J Rad Oncol Biol Phys*. 2000;47(4):1121–1135
78. Stroom JC, de Boer HC, Huizenga H, Visser AG. Inclusion of geometrical uncertainties in radiotherapy treatment planning by means of coverage probability. *Int J Radiat Oncol Biol Phys*. 1999;43:905–919.
79. Wolthaus JW, Schneider C, Sonke JJ, et al. Mid-ventilation CT scan construction from fourdimensional respiration-correlated CT scans for radiotherapy planning of lung cancer patients. *Int J Rad Oncol Biol Phys*. 2006;65:1560–1571.
80. Wolthaus JW, Sonke JJ, van Herk M, et al. Comparison of different strategies to use four-dimensional computed tomography in treatment planning for lung cancer patients. *Int J Rad Oncol Biol Phys*. 2008;70:1229–1238.
81. Liver cancer – stages. http://www.cancer.net/cancer-types/liver-cancer/stages; 2016. Accessed1:0302017-03-27
82. Shepard AJ, Wang B, Foo TK, Bednarz BP. A block matching based approach with multiple simultaneous templates for the real-time 2D ultrasound tracking of liver vessels. *Med Phys*. 2017;44:5889–5900.
83. Lachaine, M., Falco, T. Intrafractional prostate motion management with the Clarity Autoscan system. *Med Phys Int*. 2013;1:72–80.

84. Molloy JA, Chan G, Markovic A, et al. Quality assurance of us-guided external beam radiotherapy for prostate cancer: report of AAPM task group 154. *Med Phys*. 2011;38:857–871

85. Fontanarosa D, Meer S, Bloemen-van Gurp E, Stroian G, Verhaegen F. Magnitude of speed of sound aberration corrections for ultrasound image guided radiotherapy for prostate and other anatomical sites. *Med Phys*. 2012;39:5286–5292.

86. Mast TD. Empirical relationships between acoustic parameters in human soft tissues. *Acoust Res Lett Online*. 2000;1:37–42.

87. Rit S, Van Herk M, Zijp L, Sonke JJ. Quantification of the variability of diaphragm motion and implications for treatment margin construction. *Int J Radiat Oncol Biol Phys*. 2012;82:e399–e407.

88. van Herk M, Witte M, van der Geer J, Schneider C, Lebesque JV. Biologic and physical fractionation effects of random geometric errors. *Int J Radiat Oncol Biol Phys*. 2003;57:1460–1471.

89. Brix L, Ringgaard S, Sørensen TS, Poulsen PR. Three-dimensional liver motion tracking using real-time two-dimensional MRI. *Med Phys*. 2014;41:042302.

90. Davies S, Hill A, Holmes R, Halliwell M, Jackson P. Ultrasound quantitation of respiratory organ motion in the upper abdomen. *Br J Radiol*. 1994;67:1096–1102.

91. Ecclestone G, Bissonnette JP, Heath E. Experimental validation of the van Herk margin formula for lung radiation therapy. *Med Phys*. 2013;40:111721.