Robust Motion Tracking in Liver from 2D Ultrasound Images Using Supporters

Ece Ozkan · **Christine Tanner** · **Matej Kastelic** ·
Oliver Mattausch · **Maxim Makhinya** ·
Orcun Goksel

Abstract

Purpose: Effectiveness of image-guided radiation therapy with precise dose delivery depends highly on accurate target localization, which may involve motion during treatment due to, e.g., breathing and drift. Therefore, it is important to track the motion and adjust the radiation delivery accordingly. Tracking generally requires reliable target appearance and image features, whereas in ultrasound imaging acoustic shadowing and other artifacts may degrade the visibility of a target, leading to substantial tracking errors. To minimize such errors, we propose a method based on so-called *supporters*, a computer vision tracking technique. This allows us to leverage information from surrounding motion for improving robustness of motion tracking on 2D ultrasound image sequences of the liver.

Methods: Image features, potentially useful for predicting the target positions, are individually tracked and a supporter model capturing the coupling of motion between these features and the target is learned on-line. This model is then applied to predict the target position, when the target cannot be otherwise tracked reliably.

Results: The proposed method was evaluated using the Challenge on Liver Ultrasound Tracking (CLUST)-2015 dataset. Leave-one-out cross validation was performed on the training set of 24 2D image sequences of each 1-5 minutes. The method was then applied on the test set (24 2D sequences), where the results were evaluated by the challenge organizers, yielding 1.04 mm mean and 2.26 mm 95%ile tracking error for all targets. We also devised a simulation framework to emulate acoustic shadowing artifacts from the ribs, which showed effective tracking despite the shadows.

Conclusions: Results support the feasibility and demonstrate the advantages of using supporters. The proposed method improves its baseline tracker, which uses optic-flow and elliptic vessel models, and yields the state-of-the-art real-time tracking solution for the CLUST challenge.

E. Ozkan, C. Tanner, M. Kastelic, O. Mattausch, M. Makhinya, O. Goksel
Computer-assisted Applications in Medicine Group, Computer Vision Lab, ETH Zurich
E-mail: {eoezkan,tanner,ogoksel}@vision.ee.ethz.ch

1 Introduction

Ultrasound (US) imaging is a low-cost, real-time, and non-ionizing method, which
makes it an appealing choice for image-guided computer-assisted interventions in ra-
diation therapy. Treatments of liver tumors using high-intensity focused ultrasound,
intensity-modulated radiation therapy, or proton therapy enable precise dose delivery
to the desired location. However, the target region during the treatment is affected by
internal body motion, such as breathing, which is a major drawback in effectiveness
of these treatments. Not taking the respiratory motion into account would cause devi-
ations of the delivered dose distribution from the intended one, and increase radiation
exposure of healthy tissue while lowering dose to the target volume, which would
reduce efficiency and aggregate complications [1].

One of the strategies to reduce breathing-induced organ motion during radiation
treatment is deep inspiration breath hold method [2], where a patient performs a su-
pervised breath hold during therapy, which requires active support and the ability of
the patient to maintain such a breath-hold. Another possible approach to compensate
for breathing motion that does not require patient compliance, is to track the position
of the target region during therapy and dynamically adjust the radiation accordingly.

To use motion tracking algorithms for radiation therapy interventions, real-time,
accurate, and robust localization of the target region for the entire procedure is re-
quired. US imaging being non-ionizing and real-time makes it an ideal choice for
this aim [3]. There are numerous studies focusing on tracking of liver motion in US
image sequences using different approaches, such as image registration [4], block
matching [5], and optic-flow [6]. However, these methods are generally affected by
limitations of US imaging such as low signal-to-noise ratio (SNR) and large appear-
ance changes of the tracked landmarks caused by e.g. acoustic shadowing due to poor
transducer-skin contact or highly reflecting anatomical structures like the ribs.

In this work, we propose to use *supporters*, a computer vision technique [7], to
improve optic-flow based tracking. This relies on tracking additional image features,
potentially beneficial for predicting the target position. To that end, a supporter model
is built based on motion coupling observed on some frames between these tracked
features (supporters) and the target. Using this model, the tracking can then be made
robust to changes in target appearance, where a consensus voting of several supporter
estimations can be used to infer target location.

Considering motion tracking in medical images, supporters were used earlier for
determining two orthogonal MR acquisition planes through the heart valve [7]. In-
stead of the valve itself, which may leave the image, four annotated points (sup-
porters) on a plane perpendicular to the valve were tracked to define the acquisi-
tion planes. A supporter model based on squared Euclidean distances was used to
downgrade distant supporters. In [8], supporters were used for tracking abnormalities
in video capsule endoscopy. First, the supporters were matched between successive
frames by considering a triangular constraint, where the triangle shape is maintained

while allowing weak deformations. Then, affine transformations calculated from the supporter triplet help determine abnormal positions, where the precise position is estimated from the features of the target itself. In [9], cells were tracked in spatio-temporal optical images from densely packed multilayer tissues. The tight spatial topology of neighboring cells were exploited as contextual information by applying spatio-temporal graph labeling. In [10], 600 supporters were detected in fluoroscopy images by using Kanade-Lucas-Tomasi feature tracker for automatic motion compensation. An autoregression model and motion clustering was employed for learning the relationship between supporter and target motion. Supporters were also used in many other typical computer vision applications, e.g. in [11–15]. Supporters have not been studied for motion tracking in US images. We hereby show that this method is particularly beneficial in cases where the target cannot be observed directly, such as due to occlusions from shadowing artifacts.

Note that, particular challenges of US tracking are poor image quality and the relatively small number of landmarks suitable for tracking. Nevertheless, relative locations of liver landmarks stay stable during radiation therapy of liver tumors, which motivates the use of supporters in this work for 2D US tracking of the liver. We hereby devise an approach for effective supporter model creation from few supporters and evaluate this on a standard public dataset.

2 Methods

Motion tracking is the process of estimating the trajectory of an object over time by predicting its position in every frame of an image sequence. For image-guided computer-assisted applications, targets in moving organs such as the liver, prostate, and the heart are commonly tracked. Tracking an object position can be challenging, e.g. due to the appearance change over time, low SNR, or occlusions. In US images, tracked target can temporarily disappear by going out of the field-of-view or by being covered by a shadow due to poor transducer-skin contact or highly reflecting anatomical structures such as the ribs. To improve robustness of a conventional tracking algorithm for such cases, we propose combining it with a supporter model, which takes advantage of correlated surrounding motion.

2.1 Tracking with a Supporter Model

Grabner et al. [7] proposed a method for *tracking the invisible* using a set of local image features, called *supporters*, by exploiting the visual context and relative spatial relations to improve target tracking. *Good* supporters were defined as the image features whose motion are correlated to that of the target and, thus, might be useful for predicting the position of the target. For example, a wristwatch on a hand holding a target object is a good supporter for the position of that target (even when the target is not directly visible or trackable), since their motions are strongly correlated. Below we first summarize the supporter model [7] for sake of completeness and then describe our methods for its adaption in this work.

Overview of Supporter Modeling. Tracking with supporters has two main modes: learning the model and applying the model. The model captures the statistical relationship between the target and supporter positions, and thereby provides a measure of how strongly the motion between each supporter and the target is coupled. This measure can then be used for adjusting the contribution of each supporter in the overall supporter prediction.

The overall goal is to learn and apply a probability density function (pdf) model, $P(\mathbf{x}|\mathbf{I})$, for predicting the position of target object, $\mathbf{x} = (x, y)$, in image $\mathbf{I}$ via the help of S tracked supporter positions $\{\mathbf{x}_s | s = 1, 2, ..., S\}$. For this aim, the relationship between supporter positions $\{\mathbf{x}_s\}$ and the target position $\mathbf{x}$ is learned, providing conditional pdf $P(\mathbf{x}|\mathbf{x}_s)$ for supporter s. Each supporter s then votes for potential target positions $\mathbf{x}$ via pdf $P(\mathbf{x}|\mathbf{x}_s)$. These votes are combined by accounting for the reliability of the supporter position estimates $\mathbf{x}_s$ from $\mathbf{I}$ with probability $P(\mathbf{x}_s|\mathbf{I})$ using the law of total probability, which results in pdf

$$P(\mathbf{x}|\mathbf{I}) \propto \sum_{s=1}^{S} P(\mathbf{x}|\mathbf{x}_s)P(\mathbf{x}_s|\mathbf{I}). \tag{1}$$

The final target position is then determined by finding the position that has the highest likelihood in the voting space.

Learning a Supporter Model. Let $\mathbf{I}^0, \mathbf{I}^1, \ldots, \mathbf{I}^{F-1}$ be an image sequence consisting of F image frames, $\{\mathbf{x}_s^0 | s = 1, 2, ..., S\}$ be the set of S supporter positions of the first frame $\mathbf{I}^0$, and $\mathbf{x}^0$ be the target position of $\mathbf{I}^0$. The goal of the model is to estimate for frame I^f the most likely target position $\mathbf{x}^f$ from the observed supporter positions $\{\mathbf{x}_s^f\}$. Assuming a translational relationship, this is based on learning per supporter s the conditional pdf of the relative target position $\mathbf{u}_s = \mathbf{x} - \mathbf{x}_s$ for a given $\mathbf{x}_s$. For online learning during tracking, the *exponential forgetting principle* between the so-far learned pdf model $P^{f-1}(\cdot)$ and the current pdf $p(\cdot)$ is used:

$$P^f(\mathbf{u}_s|\mathbf{x}_s) = \alpha \, P^{f-1}(\mathbf{u}_s|\mathbf{x}_s) + (1-\alpha) \, p(\mathbf{u}_s^f|\mathbf{x}_s^f), \tag{2}$$

$$P^f(\mathbf{x}_s|\mathbf{I}) = \alpha \, P^{f-1}(\mathbf{x}_s|\mathbf{I}) + (1-\alpha) \, p(\mathbf{x}_s^f|\mathbf{I}^f), \tag{3}$$

where forgetting factor $\alpha \in [0, 1]$ weights the contribution of past and current pdfs. $P^f(\mathbf{u}_s|\mathbf{x}_s)$ is the model learned from frames 1 to f and provides the pdf of supporter position $\mathbf{x}_s$ voting for relative target position $\mathbf{u}_s$. $p(\mathbf{u}_s^f|\mathbf{x}_s^f)$ is the corresponding pdf derived only from the tracked positions in the current frame f. $P^f(\mathbf{x}_s|\mathbf{I})$ is the reliability model of the supporter position estimation learned from frames 1 to f. $p(\mathbf{x}_s^f|\mathbf{I}^f)$ defines the reliability of supporter position $\mathbf{x}_s^f$. We next explain how $P^f(\cdot)$ and $p(\cdot)$ are defined in practice in Section 2.2.

Applying the Supporter Model. Given image $\mathbf{I}^f$ and tracked supporter positions $\{\mathbf{x}_s^f\}$, the learned supporter models $P^f(\mathbf{u}_s|\mathbf{x}_s)$ and $P^f(\mathbf{x}_s|\mathbf{I})$ are evaluated for $\mathbf{x}_s = \mathbf{x}_s^f$ and $\mathbf{I} = \mathbf{I}^f$. From this the target position $\mathbf{x}^f$ is estimated by using Eq. (2) and Eq. (3) in Eq. (1), where the pdfs for the relative target positions are brought into the target space via $P^f(\mathbf{x} = \mathbf{u}_s + \mathbf{x}_s^f|\mathbf{x}_s^f) = P^f(\mathbf{u}_s|\mathbf{x}_s^f)$, i.e.

$$\mathbf{x}^f = \arg\max_{\mathbf{x}} P(\mathbf{x}|\mathbf{I}^f) \quad \text{with} \quad P(\mathbf{x}|\mathbf{I}^f) = \sum_{s=1}^{S} P^f(\mathbf{x}|\mathbf{x}_s^f)P^f(\mathbf{x}_s^f|\mathbf{I}^f). \tag{4}$$

2.2 Robust Motion Tracking by Estimating the Target Position Using Supporters

Tracking with supporters requires another tracking method to compute supporter locations and their reliability. Supporters can then assist and correct such a baseline method to achieve improved tracking results. We first summarize our method for a generic object tracker (see also Alg. 1), and then instantiate it with a particular tracking method later below.

Input Data. Our method uses a given initial target position $\mathbf{x}^0$, a fixed set of initial supporter positions $\{\mathbf{x}_s^0\}$, and reference patches around the target, $\mathbf{B}^0$, and each supporter, $\{\mathbf{B}_s^0\}$, where positions and reference patches are manually annotated in the first image frame $\mathbf{I}^0$. Note that the reference patches are manually chosen to contain distinct image appearance compared to their surrounding. For the current frame $f > 0$, we obtain target and supporter position estimations from the conventional object tracker, which are denoted as $\mathbf{x}_t^f$ and $\{\mathbf{x}_s^f\}$ respectively.

Tracking Reliability. Assuming that the feature appearance changes only linearly during tracking, we use the correlation coefficient measure between image patches for estimating the tracking reliability. For this, we extract patches $\mathbf{B}^f$ and $\mathbf{B}_s^f$, of the same size as $\mathbf{B}^0$ and $\mathbf{B}_s^0$, centered around the tracked positions $\mathbf{x}_t^f$ and $\mathbf{x}_s^f$, respectively. Then, we calculate the correlation coefficient between the corresponding patches, i.e. $\rho^f = CC(\mathbf{B}^0, \mathbf{B}^f)$ and $\rho_s^f = CC(\mathbf{B}_s^0, \mathbf{B}_s^f)$. We employ reliability measure ρ^f to decide whether to rely on the current target position for tracking and updating the model. Specifically, if $\rho^f \geq \theta_{CC}$, which is a learned threshold, we assume to have reliable object tracking and use this position, i.e. $\mathbf{x}^f = \mathbf{x}_t^f$. Furthermore, for another threshold $\theta_{update} > \theta_{CC}$, if $\rho^f \geq \theta_{update}$, then the supporter model is updated as described next.

Supporter Model Learning. The supporter model $P^f(\mathbf{u}_s | \mathbf{x}_s^f)$ from Eq. (2) is approximated with a 2D Gaussian distribution by

$$P^f(\mathbf{u}_s | \mathbf{x}_s^f) \propto \frac{1}{2\pi\sqrt{|\mathbf{C}_s^f|}} \exp\left(-\frac{1}{2}(\mathbf{u}_s - \boldsymbol{\mu}_s^f)(\mathbf{C}_s^f)^{-1}(\mathbf{u}_s - \boldsymbol{\mu}_s^f)^\top\right), \qquad (5)$$

Algorithm 1 Robust Motion Tracking

1: **for** each frame f **do**	12: **end if**
2: **if** $f = 0$ **then**	13: **else**
3: annotate $\mathbf{x}^0$, $\{\mathbf{x}_s^0\}$, $\mathbf{B}^0$ and $\{\mathbf{B}_s^0\}$	14: compute target probability $P(\mathbf{x}_t^f)$
4: **else**	15: **if** $P(\mathbf{x}_t^f) \geq \theta_P$ **then**
5: get $\mathbf{x}_t^f$ and $\mathbf{x}_s^f$ from object tracker	16: use object tracker: $\mathbf{x}^f = \mathbf{x}_t^f$
6: extract $\mathbf{B}^f$ and $\{\mathbf{B}_s^f\}$	17: **else**
7: compute ρ^f between $\mathbf{B}^0$ and $\mathbf{B}^f$	18: use supporter model $\mathbf{x}_p^f$: (9)
8: **if** $\rho^f \geq \theta_{CC}$ **then**	19: **end if**
9: use object tracker: $\mathbf{x}^f = \mathbf{x}_t^f$	20: **end if**
10: **if** $\rho^f \geq \theta_{update}$ **then**	21: **end if**
11: update supporter model: (6-7)	22: **end for**

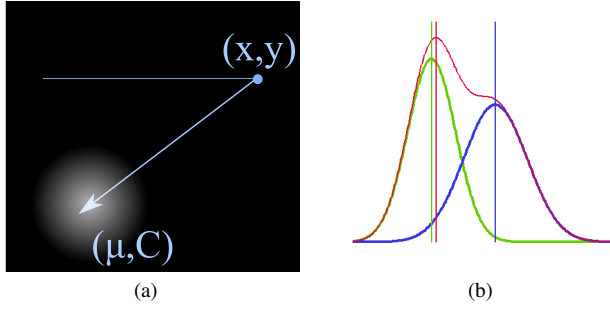(a)   (b)

Fig. 1 (a) Illustration of a supporter voting for a target position (arrow) with a probability distribution (image intensities) defined by mean $\boldsymbol{\mu}$ and covariance $\mathbf{C}_s$. (b) Illustration of a 1D Gaussian Mixture Model (red) from two individual distributions (green and blue), with mean values indicated by vertical lines.

where $\boldsymbol{\mu}_s^f$ and $\mathbf{C}_s^f$ denote the on-line learned mean and covariance matrix, respectively, of the relative target positions $\mathbf{u}_s^f$ across frames, i.e.

$$\boldsymbol{\mu}_s^f = \alpha\,\boldsymbol{\mu}_s^{f-1} + (1-\alpha)\,\mathbf{u}_s^f, \tag{6}$$

$$\mathbf{C}_s^f = \alpha\,\mathbf{C}_s^{f-1} + (1-\alpha)\,\mathbf{C}_s, \tag{7}$$

where the covariance matrix $\mathbf{C}_s$ captures the variance contribution of the current relative target position $\mathbf{u}_s^f = [u_s^f, v_s^f]$ with respect to the current mean $\boldsymbol{\mu}_s^f = [\mu_{s,u}^f, \mu_{s,v}^f]$:

$$\mathbf{C}_s = \begin{bmatrix} \left(u_s^f - \mu_{s,u}^f\right)^2 & 0 \\ 0 & \left(v_s^f - \mu_{s,v}^f\right)^2 \end{bmatrix}. \tag{8}$$

An illustration of such a distribution is shown in Fig. 1(a).

Supporter Model Application. We use the supporter model to predict the target position $\mathbf{x}^f$ if the tracked target position $\mathbf{x}_t^f$ is not reliable (i.e. $\rho^f < \theta_{CC}$). The most likely relative target location per supporter s is mean $\boldsymbol{\mu}_s^f = \arg\max_{\mathbf{u}} P^f(\mathbf{u}|\mathbf{x}_s^f)$, with corresponding probability $P^f(\boldsymbol{\mu}_s^f|\mathbf{x}_s^f) = 1/(2\pi\sqrt{|\mathbf{C}_s^f|})$. Instead of predicting the target from the peak of the resulting Gaussian Mixture Model (GMM) distribution (see Fig. 1(b) for a 1D illustration) we use a weighted average of the mean values from all mixture components [16], and incorporate the reliability of the supporter position predictions, i.e. $P^f(\mathbf{x}_s^f|\mathbf{I}^f) = \rho_s^f$. The prediction from all supporters is then

$$\mathbf{x}_p^f = \frac{\sum_s (\boldsymbol{\mu}_s^f + \mathbf{x}_s^f) P^f(\boldsymbol{\mu}_s^f|\mathbf{x}_s^f) P^f(\mathbf{x}_s^f|\mathbf{I}^f)}{\sum_s P^f(\boldsymbol{\mu}_s^f|\mathbf{x}_s^f) P^f(\mathbf{x}_s^f|\mathbf{I}^f)} = \frac{\sum_s (\boldsymbol{\mu}_s^f + \mathbf{x}_s^f)\rho_s^f/\sqrt{|\mathbf{C}_s^f|}}{\sum_s \rho_s^f/\sqrt{|\mathbf{C}_s^f|}}. \tag{9}$$

Finally, if the applied supporter model and the main object tracker agree on the target position estimation, i.e. $P(\mathbf{x}_t^f) = \sum_s P^f(\mathbf{x}_t^f - \mathbf{x}_s^f|\mathbf{x}_s)\rho_s^f \geq \theta_P$, then the estimation from the main tracker is used: $\mathbf{x}^f = \mathbf{x}_t^f$. Otherwise, we use the supporter prediction $\mathbf{x}^f = \mathbf{x}_p^f$. An example for target position estimation using supporter model is shown in Fig. 2.
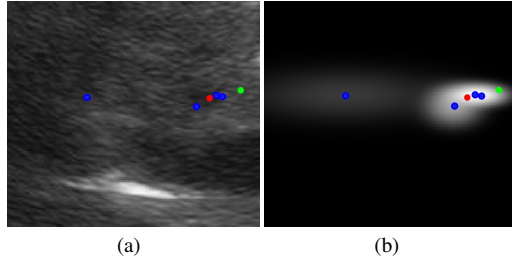
(a) (b)

Fig. 2 Example of tracker and supporter predictions. Target position from main object tracker $\mathbf{x}_t^f$ (green), individual supporter predictions $\boldsymbol{\mu}_s^f + \mathbf{x}_s^f$ (blue) and weighted mean using Gaussian Mixture Model $\mathbf{x}_p^f$ (red) overlaid on (a) US image and (b) log transformed probability density.

3 Experiments and Results

We evaluated our method using the 2D liver US image sequences provided by the Challenge on Liver Ultrasound Tracking (CLUST)-2015 [17]. A main advantage of supporters is the robustness to feature appearance in tracking, for instance, when a target is occluded by acoustic shadowing. Since such disappearing target locations are not (and cannot reliably be) annotated in the given dataset, we devised a simulation framework to emulate acoustic shadowing artifacts from the ribs on the images and evaluated this scenario. As the baseline object tracker, we employed [6] such that motion tracking with and without using the supporter model can be compared.

3.1 CLUST-2015 Dataset

The CLUST-2015 dataset includes 2D liver US image sequences and consists of two subsets, namely training and test set. The sequences in the dataset have a duration between 60 and 330 seconds. The training set has 24 image sequences with manual annotations in 10% of all frames. The annotations are mostly for vessel cross-sections in the liver, which are reliable landmarks for liver motion. The test set contains 24 image sequences with no public annotations apart from the reference positions $\mathbf{x}^0$, and the submitted results are evaluated by the challenge organizers. For the evaluation, the Euclidean distance between each manual annotation and the corresponding tracked point is computed, where summary error statistics including mean, standard deviation, and 95%ile errors are reported to the participant. In this work, we are particularly interested in reducing 95%ile errors to minimize large errors for a robust tracking performance throughout all sequences.

For parameter optimization and sensitivity analysis, we used the training set. Our method has four parameters to optimize, which are forgetting factor α, correlation coefficient threshold θ_{CC}, supporter model update threshold θ_{update}, and target probability threshold θ_P. We optimized these parameters for minimizing 95%ile error with leave-one-out-cross validation using grid search. Optimal parameters range from $[\alpha, \theta_{CC}, \theta_{update}, \theta_P] = [0.90, 0.3, 0.3, 0.5]$ to $[0.95, 0.3, 0.4, 0.7]$ and hence are relatively insensitive to the left-out case. The mean parameters were found to be
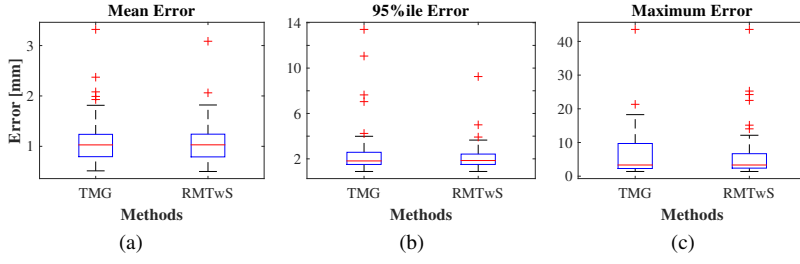
Fig. 3 Tracking error distributions [in mm] for baseline (TMG) and proposed method (RMTwS) for **24 training sequences**. (a) Mean tracking error. (b) 95%ile tracking error. (c) Maximum tracking error.

Overall Performance for Training and Test Set

Method	Training Set				Test Set			
	Mean	σ	95%ile	Max	Mean	σ	95%ile	Max
TMG	1.17	0.89	2.61	21.78	1.09	1.75	2.42	25.55
RMTwS	1.12	0.81	2.19	21.78	1.04	1.48	2.26	21.41
	(a)				(b)			

Table 1 Comparison of mean, standard deviation, 95%ile and maximum of tracking errors (in mm) of baseline (TMG) and proposed (RMTwS) method after pooling all results from (a) training and (b) test set.

$[\alpha, \theta_{CC}, \theta_{update}, \theta_P] = [0.9479, 0.3000, 0.3021, 0.6625]$. Fig. 3 shows the mean, 95%ile, and maximum tracking error distributions from the 24 sequences of the baseline method (abbreviated as *TMG* for *Tracking by Makhinya and Goksel*) and our proposed tracker (denoted as *RMTwS* for *Robust Motion Tracking with Supporters*). Table 1a compares overall performance for the mean, standard deviation, 95%ile, and maximum error after pooling all training results into one distribution. Note that our proposed method yields a 16% improvement for the 95%ile error. Average error of the worst 5% tracking results across all training annotations is 10.70 mm with TMG. Our proposed technique yielded 4.7 mm, improving the baseline over 50%.

We then applied our method on the test set using the optimal parameters found above. Test set results were evaluated by the challenge organizers. Fig. 4 compares tracking error distributions of the baseline tracker, TMG, and our proposed tracker, RMTwS, for the 24 test sequences, and Table 1b lists the overall performance after pooling all results. RMTwS yields 1.04 mm mean and 2.26 mm 95%ile error, improving the baseline method by 4.6% and 6.6%, respectively. The 95%ile error of the 62 individual test landmarks was improved by more than 30% for 5 landmarks, whereas for 55 landmarks the improvements were less than 2%.

We also evaluated the time needed to run our proposed method. Learning and applying the supporter model takes between 20 and 60 ms per frame in the given sequences on an Intel Core i7-4770K CPU @ 3.5GHz.

3.2 Evaluating Tracking under Shadowing

Since the target points which disappear in the acoustic shadow are not annotated in the CLUST-2015 dataset, we conducted a simulation, where we emulated acoustic
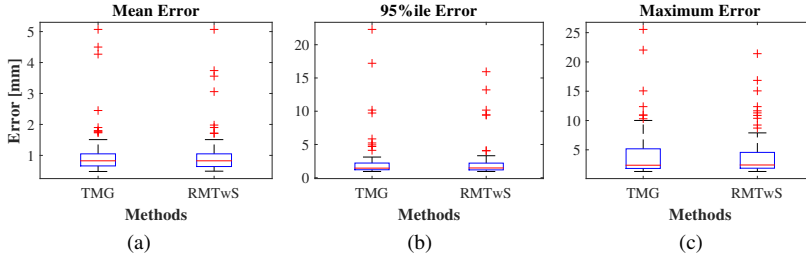
Fig. 4 Tracking error distribution [in mm] for baseline (TMG) and proposed method (RMTwS) for **24 test sequences**. (a) Mean tracking error. (b) 95%ile tracking error. (c) Maximum tracking error.

shadowing artifacts from a simulated rib on the images and evaluated this scenario. For this purpose, we manually placed a structure of size 12.4 mm × 7.2 mm, representing a rib cross-section in accordance with [18], close to the skin.

We augmented each frame in a US image sequence from the training data with new ultrasound bone shadows by multiplying the input US images with a *signal intensity map*. For each pixel of an ultrasound image, this map stores the accumulated intensity of the ultrasound signal induced by reflection at the bone surface and energy loss (attenuation) within the bone structures. It is between $[0,1]$, with 1 for the original signal intensity and 0 for a complete signal loss. The signal intensity map is generated in a multi-stage process. In the first step, we create a map of attenuation coefficients $\mathbf{Z}$ of bone cross sections, given by intersection of the bone tissue with the transducer plane. To create a bone segment j, we simply rasterize a circle with radius r_j at position p_j in $\mathbf{Z}$. Inside each circle, we store attenuation coefficients $\mathbf{Z}(x,y) = \beta_j$ corresponding to bone segment j, and $\mathbf{Z}(x,y)$ is zero otherwise. Typical values of β for bone are used from literature [19].

In the next step, we use *ray marching* to traverse $\mathbf{Z}$ and create a (pre-scan-converted) signal intensity map $\mathbf{A}$, in a simplified and task-specific variation of more complex ultrasound simulation method [19]. In particular, we traverse the columns (scanlines) of $\mathbf{Z}$ from top to bottom (y-direction). During this, we record a reflected signal intensity at the bone surface and energy loss thereafter, and accumulate the attenuation coefficients in $\mathbf{Z}$. At each step of the ray marching process, the current pixel $\mathbf{A}(x,y)$ is computed as $\mathbf{A}(x,y) = \mathbf{A}(x,y-1)\exp(-\mathbf{Z}(x,y))$.

The resulting signal intensity map is finally filtered with a Gaussian function to emulate the blurring due to convolution with the ultrasound point spread function (PSF). Since the input images are from a convex probe, the map is scan-converted from a radial domain into a Cartesian frame, using the the scan-conversion parameters estimated geometrically from the original image. This yields the typical ultrasound shadow appearance in convex probe images, where the shadows become softer and wider in the far field of the images. This provides simulated image data with ground-truth for evaluating tracking under shadowing. Example images of a signal intensity map, an original image, and the resulting shadowed image are shown in Fig. 5.

After generating a 2D US image sequence containing shadow, we applied the baseline and our method to the new sequence. For that, we used the same optimal pa-
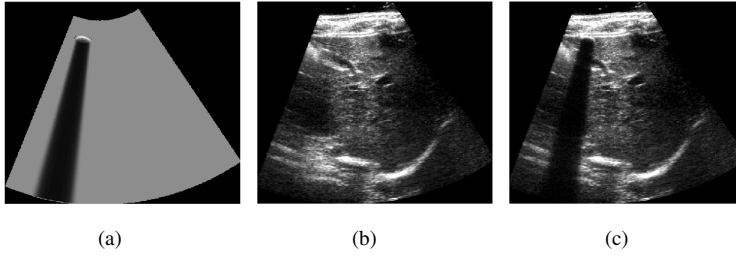
Fig. 5 Shadow simulation example with (a) signal intensity map, (b) original image, (c) shadowed image.
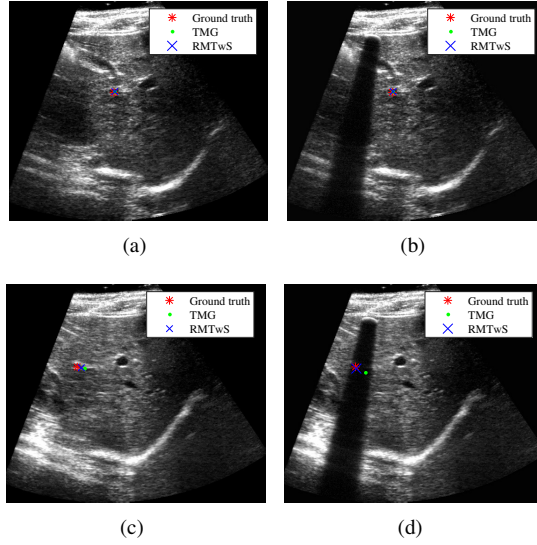

Fig. 6 Example of tracking performance (a)(c) without and (b)(d) with shadowing for (a)(b) inhale and (c)(d) exhale breathing phase, showing improved robustness of the proposed method, *RMTwS*, in (d).

rameters as for the CLUST-2015 test set, obtained by leave-one-out-cross validation. The mean errors for TMG and RMTwS were 2.79 mm and 2.61 mm, with 95%ile errors of 12.11 mm and 10.29 mm. This indicates a 6.5% (15%) improvement of mean (95%ile) error. Examples of tracking performance with and without shadowing for inhale and exhale phases of the breathing cycle are shown in Fig. 6.

4 Discussion and Conclusions

We have demonstrated an ultrasound tracking method using supporters, RMTwS, where image locations other than the target are also tracked in order to exploit motion consistency with such surrounding tissue for improving tracking robustness. We employed an optic-flow and vessel-model based tracker, TMG, as our baseline as well as for tracking the target and supporter locations to then learn and apply the supporter

model using these initial estimations. In this work, we are particularly interested in reducing 95%ile errors to ensure effective tracking performance throughout all frames in order to minimize 95%ile therapy margins for more focal therapies and reduced collateral damage to healthy tissue.

Our evaluations using the training and test sets show that the proposed method, RMTwS, can track targets more accurately than the conventional object tracker, TMG. The resulting performance is 1.04 mm mean and 2.26 mm 95%ile errors. This 95%ile tracking performance is relevant in liver motion tracking for radiation and focused therapy applications, when compared to 1.23 mm mean inter-observer 95%tile variability reported for a similar dataset in [17].

The accuracy improvements seem to be small for mean and 95%ile error, when taking all trajectories into account. This is because the main object tracker already performs quite well in most cases and fails only in certain situations such as under shadowing. All the same, to enable a satisfactory therapy for every patient, a tracking method should be robust for all scenarios.

Optimal thresholds for updating the supporter model, θ_{update}, and the reliability of the tracking performance, θ_{CC}, were found to be very close. A supplementary experiment showed that the tracking performance difference using $\theta_{update} = \theta_{CC}$ is insignificant. Thus, one can use the same parameter for θ_{CC} and θ_{update}.

Our proposed method applies the learned supporter model in 12% of the frames, which indicates that the reliability of the tracking performance by TMG is not always high. The main advantage of using supporters for tracking is the robustness in scene or target appearance changes over time, such as due to acoustic shadowing. Since there exist no annotations for such cases in the given dataset and this scenario cannot be evaluated using the current setting, we devised a simulation framework to imitate acoustic shadowing artifacts on the images in a 2D sequence. This simulated experiment showed that without additional optimization for such a scenario, the proposed method improves the 95%ile tracking performance of the baseline by 15%.

On each sequence 2 to 3 supporters were used, which is not a large number since there are only a few easily identifiable landmarks in these images. We aim to study automatic landmark detection in the future to automatically identify a (potentially larger) number of supporters, also yielding a interaction-free framework. Additionally, with more supporters available, we plan to conduct a sensitivity analysis regarding their number and locations.

We currently use the Gaussian position prior model with individual translational motion assumptions between the target and each supporter. Nevertheless, the combination of several supporter estimates can indeed lead to relative positions that are not purely translation between supporter locations and the target. Although more complex relative (e.g., elastic) motion models could be employed, these could, however, complicate extrapolating target locations further away from the supporters.

This study is the first demonstrating the benefits of employing supporters for US tracking. Given the target and supporter position estimations from the main object tracker, learning and applying the supporter model takes less than 20 ms, where correlation coefficient calculation takes most of it. The resulting tracking technique has a near real-time tracking performance with 22.5 frames per second (fps) on average. As such, it is the state-of-the-art in the CLUST2015 challenge for real-time tracking

of liver motion in 2D ultrasound sequences, as the winner of this challenge achieved a mean (95%ile) error of 0.91 mm (2.20 mm) while running on average at 4.8 fps; hence, our method being more than 4 times faster in comparison.

In a practical application of our method in radiation therapy, a 2D convex transducer can be used to image the liver reaching below the ribs. On an initial (reference) frame, an operator would then mark the target location, as well as a few other easy-to-track locations (supporters). Tracking would then run during the treatment, while the target location estimates are used to gate or compensate for patient motion.

Appendix: Baseline Method - *T*racking by *M*akhinya and *G*oksel (*TMG*)

Our previously developed tracker [6], which is runner up of the Challenge in Liver Ultrasound Tracking (CLUST)-2015 challenge and is based on optic-flow and elliptic vessel model, is employed as object tracker for tracking the supporters and target. The method is summarized below for completeness. Note that, this method can track several landmarks together real-time and works faster than US acquisition.

Overview The method decides in the initial frame, if the target is vessel-like or not by matching with ellipsoid vessel templates and integrates then several tracking strategies. It involves reference tracking (RT) when the local appearance on the initial, $\mathbf{I}^0$, and the current frame, $\mathbf{I}^f$, are similar. Meanwhile, it uses model-based iterative tracking (IT) when RT fails and local appearance of consecutive frames, $\mathbf{I}^{f-1}$ and $\mathbf{I}^f$, are similar. A robust *motion tracking* is applied in either case. For vessel-like structures this is improved further by *model-based tracking*.

Motion Tracking Lucas-Kanade-based tracking [20] was applied on a set of regularly-spaced grid points around each target. RT is then used for exploiting the repetitive breathing motion characteristic, while IT is used for tracking the motion during the rest of the cycle, i.e. when RT fails. Each tracking strategy yields several motion vectors, which are then filtered for outliers. Finally, from the remaining motion vectors, an affine transform is computed to provide a robust motion estimate for the target.

Model-based Tracking For vessel-like structures, model-based tracking is done using an axis-aligned ellipse representation of vessels. For each frame $\mathbf{I}^f$, first the center is transformed by the affine transform determined by motion tracking, see above, and then the center and radii are re-estimated as in [21] using the Star Edge detection, dynamic programming, model fitting, and binary template matching. The center of the resulting ellipse is then used as the estimated target position at frame $\mathbf{I}^f$.

References

1. P. J. Keall, G. S. Mageras, J. M. Balter, R. S. Emery, K. M. Forster, S. B. Jiang, J. M. Kapatoes, D. A. Low, M. J. Murphy, B. R. Murray, C. Ramsey, M. Van Herk, S. Vedam, and E. Wong, J.W. Yorke, "The management of respiratory motion in radiation oncology report of AAPM Task Group 76a)," *Medical Physics*, vol. 33, no. 10, pp. 3874–3900, 2006.

2. G. S. Magerasa and E. Yorkea, "Deep inspiration breath hold and respiratory gating strategies for reducing organ motion in radiation treatment," *Seminars in Radiation Oncology*, vol. 14, no. 1, pp. 65–75, 2004.

3. V. De Luca, G. Szkely, and C. Tanner, "Estimation of large-scale organ motion in B-mode ultrasound image sequences: A survey," *Ultrasound in Medicine and Biology*, vol. 41, no. 12, pp. 3044–3062, 2015.

4. S. Vijayan, S. Klein, E. F. Hofstad, F. Lindseth, B. Ystgaard, and T. Langø, "Validation of a non-rigid registration method for motion compensation in 4D ultrasound of the liver," in *2013 IEEE 10th International Symposium on Biomedical Imaging*, pp. 792–795, 2013.

5. V. De Luca, M. Tschannen, G. Székely, and C. Tanner, "A learning-based approach for fast and robust vessel tracking in long ultrasound sequences," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 518–525, Springer, 2013.

6. M. Makhinya and O. Goksel, "Motion tracking in 2D ultrasound using vessel models and robust optic-flow," in *MICCAI 2015 Challenge on Liver Ultrasound Tracking*, 2015.

7. H. Grabner, J. Matas, L. Van Gool, and P. Cattin, "Tracking the invisible: Learning where the object might be," in *International Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1285–1292, 2010.

8. Y. Yanagawa, T. Echigo, H. Vu, H. Okazaki, Y. Fujiwara, T. Arakawa, and Y. Yagi, "Tracking abnormalities in video capsule endoscopy using surrounding features with a triangular constraint," in *International Symposium on Biomedical Imaging (ISBI)*, 2012.

9. A. Chakraborty and A. K. Roy-Chowdhury, "Context aware spatio-temporal cell tracking in densely packed multilayer tissues," *Medical Image Analysis*, vol. 19, no. 1, pp. 149–163, 2015.

10. Y. Xia, S. Hussein, V. Singh, M. John, Y. Wu, and T. Chen, "Context region discovery for automatic motion compensation in fluoroscopy," *International Jurnal of Computer Assisted Radiology and Surgery*, vol. 11, no. 6, pp. 1–9, 2016.

11. Z. Sun, H. Yao, S. Zhang, and X. Sun, "Robust visual tracking via context objects computing,"," in *18th IEEE International Conference on Image Processing*, pp. 509–512, 2011.

12. F. Xiong, O. I. Camps, and M. Sznaier, "Dynamic context for tracking behind occlusions," in *European Conference on Computer Vision (ECCV)*, pp. 580–593, 2012.

13. L. Meng and Q. Jia, "Multi-target tracking based on level set segmentation and contextual information," *International Journal of Signal Processing, Image Processing and Pattern Recognition*, vol. 6, no. 4, pp. 287–296, 2013.

14. L. Zhang and L. Van Der Maaten, "Preserving structure in model-free tracking," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 4, pp. 756–769, 2014.

15. K. Meshgi, S.-i. Maeda, S. Oba, H. Skibbe, Y.-z. Li, and S. Ishii, "An occlusion-aware particle filter tracker to handle complex and persistent occlusions," *Computer Vision and Image Understanding*, vol. 150, pp. 81–94, 2016.

16. G. Samei, G. Chlebus, G. Sz ekely, and C. Tanner, "Adaptive confidence regions of motion predictions from population exemplar models.," in *MICCAI Workshop on Computational and Clinical Challenges in Abdominal Imaging*, pp. 231–240, 2013.

17. T. De Luca, V. annd Benz, S. Kondo, L. Knig, D. Lbke, S. Rothlbbers, O. Somphone, S. Allaire, M. Lediju Bell, D. Chung, A. Cifor, C. Grozea, M. Gnther, J. Jenne, T. Kipshagen, M. Kowarschik, N. Navab, J. Rhaak, J. Schwaab, and C. Tanner, "The 2014 liver ultrasound tracking benchmark," *Physics in Medicine and Biology*, vol. 60, no. 14, p. 5571, 2015.

18. M. Mohr, E. Abrams, C. Engel, W. B. Long, and M. Bottlang, "Geometry of human ribs pertinent to orthopedic chest-wall reconstruction," *Journal of Biomechanics*, vol. 40, no. 6, pp. 1310–1317, 2007.

19. O. Mattausch and O. Goksel, "Monte-carlo ray-tracing for realistic interactive ultrasound simulation," in *Eurographics Workshop on Visual Computing for Biology and Medicine*, 2016.

20. B. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in *Proceedings of Imaging Understanding Workshop,*, pp. 121–130, 1981.

21. A. Crimi, M. Makhinya, U. Baumann, C. Thalhammer, G. Szekely, and O. Goksel, "Automatic measurement of venous pressure using B-mode ultrasound," *IEEE Transaction on Biomedical Engineering*, vol. 63, no. 2, pp. 288–299, 2016.