Model-based Sparse-to-dense Image Registration for Realtime Respiratory Motion Estimation in Image-guided Interventions

In Young Ha*, Matthias Wilms*, Heinz Handels, and Mattias P. Heinrich

Abstract*—*Objective: **Intra-interventional respiratory motion estimation is becoming a vital component in modern radiation therapy delivery or high intensity focused ultrasound systems. The treatment quality could tremendously benefit from more accurate dose delivery using realtime motion tracking based on magnetic-resonance (MR) or ultrasound (US) imaging techniques. However, current practice often relies on indirect measurements of external breathing indicators, which has an inherently limited accuracy. In this work, we present a new approach that is applicable to challenging realtime capable imaging modalities like MR-Linac scanners and 3D-US by employing contrast-invariant feature descriptors.** ***Methods:*** **We combine GPU-accelerated image-based realtime tracking of sparsely distributed feature points and a dense patient-specific motion-model for regularisation and sparse-to-dense interpolation within a unified optimisation framework.** ***Results:*** **We achieve highly accurate motion predictions with landmark errors of ≈1 mm for MRI (and ≈2 mm for US) and substantial improvements over classical template tracking strategies.** ***Conclusion:*** **Our technique can model physiological respiratory motion more realistically and deals particularly well with the sliding of lungs against the rib cage.** ***Significance:*** **Our model-based sparse-to-dense image registration approach allows for accurate and realtime respiratory motion tracking in image-guided interventions.**

***Index Terms*—respiratory motion estimation, sparse-to-dense registration, MRI-guided interventions, HIFU**

I. Introduction

Respiratory motion during the course of non-invasive image-guided radiation therapy (IGRT) using a linear accelerator (Linac) or high intensity focused ultrasound (HIFU) can have a negative impact on treatment qualities and should be addressed by intra-interventional motion tracking to reduce treatment margins and potentially enable more targeted radiation delivery [1], [2]. Previously, most realtime intra-interventional respiratory motion tracking methods relied on ionising X-ray imaging [3] or indirect measurements of external breathing indicators [4]. However, the recent introduction of realtime 4D (3D+t) ultrasound (US) and magnetic resonance imaging (MRI) techniques into radiation therapy and HIFU has paved the way for more accurate and reliable intra-interventional motion tracking [5]–[7].

Whereas algorithms for deformable image registration for offline respiratory motion estimation from, e.g., CT data have reached intra-observer accuracy even for large motion amplitudes [8], they are computationally complex and cannot meet the requirement of realtime applications, where several frames have to be processed every second. Furthermore, intra-interventional motion estimation based on MRI or US data leads to at least two additional challenges that are usually not addressed by classical image registration algorithms developed for volumetric CT data: (1) A high level of image noise and spatially varying contrast, (2) a difference in field of view or image dimension between the pre-treatment reference image and the intra-interventional target image.

Whereas the first problem can be effectively tackled by using contrast-invariant feature descriptors [9] for image matching, the second problem is much more difficult to solve. For example, in US-based guidance, the field of view changes rapidly depending on the probe position/organ motion or bone shadows. Integrated MR-Linac systems will only provide few (≈5) 2D slices in sufficient temporal resolution (3-8 Hz) [10]. However, image-guided dose delivery and replanning requires the estimation of dense 3D motion of tumours and organs at risk depicted in the 3D pre-treatment planning image using this incomplete or sparse imaging information.

A. Related Work on Realtime Motion Tracking

In this section, we will review related work on realtime respiratory motion tracking in US- or MRI-guided interventions.

The most basic approach for realtime motion estimation based on intra-interventional image data is to track only few relevant locations close to the structure of interest using template/block-matching techniques [11]–[19]. These approaches are fast and reasonably accurate for the feature points tracked, but do not produce a dense motion field of the whole treatment area and, therefore, do not estimate the motion of organs at risk or other nearby structures. Furthermore, basic template/block-matching is an inherently local technique lacking a dedicated regularisation scheme to guarantee spatial regularity and/or temporal smoothness and is, hence, prone to ambiguities during the matching process.

Estimation or interpolation of dense motion fields based on sparse correspondences obtained via template/block-matching has been an active field of research lately. Techniques can be roughly divided into two categories: (1) Approaches that rely on general, unspecific interpolation techniques like Thin-Plate Splines [20] or piece-wise affine warps [21]. (2) Approaches that employ statistical motion models [22], which can be generated from pre-treatment data and subsequently be used for

This work was supported by the German Research Foundation (DFG) under grant 286491894 (HE 7364/1-2).

All authors are with the Institute of Medical Informatics, University of Lübeck, Lübeck, Germany (correspondence e-mail: ha@imi.uni-luebeck.de)

* I.Y. Ha and M. Wilms contributed equally to this work.

sparse-to-dense interpolation during the treatment [10], [23]–[26]. In contrast to unspecific techniques, (patient-specific) motion models have usually better reconstruction capabilities and are able to accurately reconstruct local patient-specific deformation details in areas with very sparse correspondences. Moreover, those models can also serve as a motion prior during the fitting process [10], [26], [27], but this aspect has not been explored much in template/block-matching-based approaches.

In addition to template/block-matching-based approaches, sparsified or otherwise accelerated versions of dense registration techniques like the Demons algorithm [28], advanced optical flow techniques [27], [29], [30], or variational image registration approaches [31] have been proposed. Although being very fast ([29]: < 50 ms for 2D-2D registration), accurate, and robust by using classical regularisers, these approaches typically cannot handle incomplete image data (e.g., few 2D slices in MRI-guided RT). Furthermore, all of them perform gradient descent-based optimisation and are, hence, prone to local minima. This is also true for the approach presented in [21], which combines intensity-based mesh model fitting with mechanical-based regularisation and which is currently the best performing algorithm on the CLUST 2014 and CLUST 2015 3D ultrasound tracking challenge data sets [18].

B. Contributions

Because of its inherent ability to handle the sparse-to-dense registration problem and its computational simplicity, we argue that tracking of sparse feature points using block-matching combined with motion model-based dense field reconstruction is the most promising technique for realtime respiratory motion estimation in MRI- and US-guided interventions. But, we strongly believe that in contrast to previous work (e.g., [23], [24]), the motion model should be incorporated into the sparse motion estimation step as a spatial prior/regulariser to improve matching robustness and accuracy. We, therefore, combine **GPU-accelerated block-matching** for tracking of sparsely distributed feature points covering both tumour and organs-at-risk, a **dense patient-specific motion-model for regularisation** and sparse-to-dense interpolation **within a unified discrete optimisation framework**. Using appropriate patient-specific motion models enables us to compute highly accurate dense displacement fields with a natural handling of sliding motion.

To our knowledge, this is the first realtime respiratory motion estimation approach for image-guided interventions based on computationally fast discrete optimisation techniques that jointly optimises sparse block-matching and model-based regularisation. Previous approaches with motion model-based regularisation, can either only handle dense image data [27], rely on continuous optimisation [10], [26], [27], or are not realtime capable [10], [26].

This work is based on our MICCAI 2016 conference paper [32] and extends it in several ways: (1) To overcome the computational bottleneck of the block-matching step presented in [32], we propose an elegant convolution-based GPU implementation and reduce processing times by over an order of magnitude. (2) We greatly extend the analysis of parameter settings. (3) We add further publicly available MR image sequences [33] to our evaluation and provide manual landmarks for this data. (4) We directly compare our approach to the current state-of-the-art for MRI based tracking, which employs advanced optical flow techniques for tracking a target in MRI [30] and can also include regional information using a PCA motion model [27]. (5) We present a landmark-based evaluation of our approach on publicly available 4D US data.

II. Materials and Methods

The general workflow of our framework is illustrated in Fig. 1 and consists of two distinct phases: In the **pre-treatment phase**, a sequence of n-dimensional images of the patient under free-breathing for T time frames $\{I_{j^*}\}_{j^*\in\{1,\dots,T\}}$ is acquired for treatment planning, target delineation and dose calculation (*training images*). From these images, we select a reference image $I_R \in \{I_{j*}\} : \Omega \to \mathbb{R}$, with the image domain Ω indicating the region of interest to be tracked during the intervention. Furthermore, these training images are used to build a patient-specific motion model (Section II-B).

During the **treatment phase** intra-interventional images $I_{M,t} : \Omega' \to \mathbb{R}$ at different time points t are acquired in realtime. Please note that the domain Ω' of the intra-interventional images is usually smaller (e.g., few image slices) and/or only a subset of the domain of the reference image ($\Omega' \subseteq \Omega$) and, therefore, a dense transformation $\varphi_t = Id + \mathbf{u}_t : \Omega \to \Omega$, with a dense displacement field $\mathbf{u}_t : \Omega \to \mathbb{R}^n$, that describes the motion of the region of interest between the reference image and time point t cannot be determined directly. For this reason and to meet realtime requirements, we opt to track a set of N sparse feature points in the reference image $\Omega'_N = \{\mathbf{x}_1, ..., \mathbf{x}_N\}$ ($\Omega'_N \subset \Omega'$) via block-matching (Section II-A).

Given the feature points, our goal is to find an optimal sparse displacement field $\tilde{\mathbf{u}}_t : \Omega'_N \to \mathbb{R}^n$ at the time point t, which minimises the joint cost function,

$$E(\tilde{\mathbf{u}}_t) = \sum_{\Omega_N} \mathcal{D}(I_R, I_{M,t}, \tilde{\mathbf{u}}_t) + \alpha\mathcal{R}(\tilde{\mathbf{u}}_t) \qquad (1)$$

where $\mathcal{D}$ is a point-wise dissimilarity measure between $I_R(\mathbf{x}_i)$ and $I_{M,t}(\mathbf{x}_i + \tilde{\mathbf{u}}_t(\mathbf{x}_i))$ around the feature point $\mathbf{x}_i$ and $\mathcal{R}$ is a regularisation term controlled by a weight parameter α. Our patient-specific motion model is used to regularise and penalise deviations of the sparse motion field $\tilde{\mathbf{u}}_t$ from plausible solutions and is also used to finally reconstruct the dense field $\mathbf{u}_t$ from $\tilde{\mathbf{u}}_t$ (Section II-B).

Due to its non-linearity the cost function in (1) is difficult to minimise. We, therefore, propose a coupled convex optimisation approach, which efficiently determines an optimal sparse field by alternately optimising over the local image dissimilarity distribution determined via block-matching and global model-based regularisation (Section II-C).

A. Feature Points Selection and Similarity-driven Block-Matching

The feature points Ω'_N can be extracted using any automatic landmark detection algorithm, which returns distinguishable

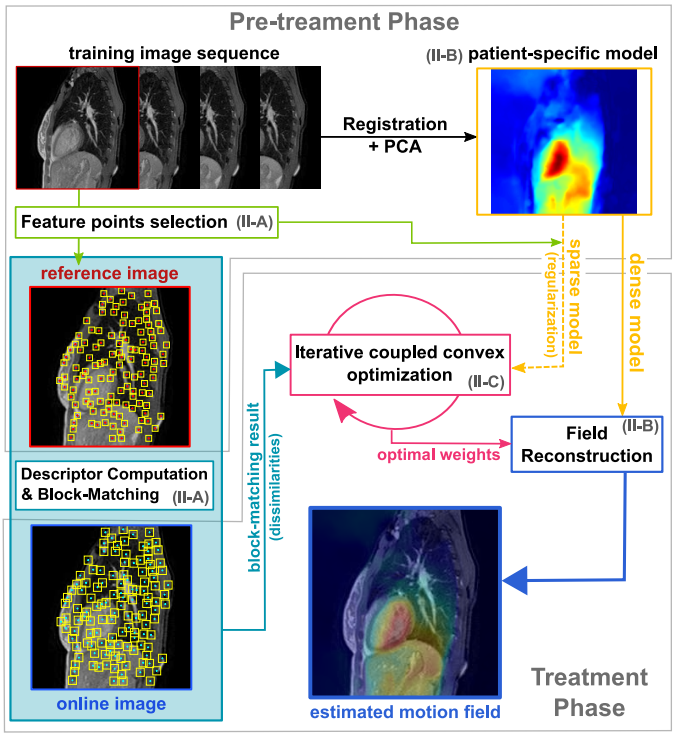

Fig. 1. Workflow of our framework. Our approach estimates realtime dense motion field by combining the patient-specific respiratory motion (Section II-B) with the sparse online motion observation (Section II-A) using the coupled convex optimisation (Section II-C).

points. In this work, we use the Harris corner detector to extract feature points within the reference image of the **pre-treatment phase**, and employ a non-maximum suppression for a good distribution of the feature points [34]. The number of selected feature points is dependent on the radius of the non-maximum suppression and a suitable choice is determined empirically.

The dissimilarity cost $\mathcal{D}$ between these feature locations in the reference image I_R and for each point $\mathbf{d}_{i,t}$ in the search region of a target image $I_{M,t}$ (acquired during the **treatment phase**) are computed using the block-matching algorithm. Given a reference image block around the feature point considered, the block-matching algorithm computes the block-wise (dis)similarity $\mathcal{D}$ between the reference block and all blocks within the search region of the target image. This results in a dissimilarity map for each feature point, which is stored for later use (Section II-C). Block-sizes will be given in the experimental section. Instead of performing the block-matching directly with the image patches which might have noise and intensity variations, we use self-similarity context (SSC) descriptors [9], which are insensitive to image intensities and noise. The initial sparse motion field $\tilde{\mathbf{u}}_t$ is also determined using the block-matching algorithm by finding the sparse displacements of minimal dissimilarity cost.

Since the block-matching algorithm optimises only the local image similarities, the initial sparse motion field $\tilde{\mathbf{u}}_t$ determined based on the block-matching result will usually contain displacement vectors inconsistent with the global motion. To find the optimal sparse motion field, which can properly describe both the local deformation and the global motion, we incorporate the patient-specific motion using model-based regularisation.

B. Patient-specific Respiratory Motion Model

Based on the training images $\{I_j\}_{j\in\{1,\dots,T\}}$, a patient-specific respiratory motion model is built by non-linearly registering each training image to the reference image I_R followed by performing a PCA on the estimated displacement fields. Although any non-linear registration algorithm can be used for registration, we chose the publicly available *deeds* algorithm, as it can handle sliding motion correctly and has shown high accuracy in respiratory motion estimations tasks [35]. As registration output, *deeds* produces a dense displacement field $\mathbf{u}_j$ for each image I_j, which encodes the respiration-related motion between I_j and I_R. To perform linear statistics on them, the displacement fields are first vectorized and then concatenated to form a data matrix $\mathbf{U} \in \mathbb{R}^{nV\times T}$, where V denotes the number of image voxels and each column vector corresponds to a displacement field.

Based on $\mathbf{U}$, principle component analysis (PCA) can be performed by doing an eigendecomposition of its sample covariance matrix:

$$\mathbf{C} = \frac{1}{T}\sum_{t=1}^{T}(\mathbf{u}_j-\mu)(\mathbf{u}_j-\mu)^T = \mathbf{P}\mathbf{\Sigma}\mathbf{P}^T \ , \qquad (2)$$

where $\mu = \frac{1}{T}\sum_{j=1}^{T}\mathbf{u}_j$ is the mean displacement field.

From the eigendecomposition, we obtain a matrix $\mathbf{P} \in \mathbb{R}^{nV\times nV}$ whose orthonormal columns are eigenvectors of $\mathbf{C}$ and span a linear space. Furthermore, diagonal matrix $\mathbf{\Sigma} \in \mathbb{R}^{nV\times nV}$ contains the corresponding eigenvalues, which represent the variance covered by each eigenvector. To end up with a compact model and for elimination of noise or other irrelevant information, typically only the eigenvectors corresponding to the k largest eigenvalues are retained where k is chosen based on a user-specified variance threshold (e.g., 95% variance to be retained) [22]–[25]. The columns of the reduced matrix $\mathbf{P}_k \in \mathbb{R}^{nV\times k}$ form an orthonormal basis, which parameterises the subspace of plausible mean-centred displacement fields (model space). Uncentred displacement fields $\mathbf{u}$ belonging to this model space can be generated via

$$\mathbf{u} = \mu + \mathbf{P}_k\mathbf{\Sigma}_k^{\frac{1}{2}}\mathbf{c} \ , \qquad (3)$$

where $\mathbf{c} \in \mathbb{R}^{k\times 1}$ is a weight vector.

Given a sparse displacement field $\tilde{\mathbf{u}}_t$ (see Section II-A), a dense displacement field $\mathbf{u}_t$ can be reconstructed, by optimally projecting $\tilde{\mathbf{u}}_t$ into model space in a regularised least-squares way by minimising

$$E(\mathbf{c}) = ||\tilde{\mathbf{P}}_k\mathbf{\Sigma}_k\mathbf{c} - (\tilde{\mathbf{u}}-\tilde{\mu})||_2^2 + \eta||\mathbf{c}||_2^2 \ , \qquad (4)$$

and plugging the estimated weights back into (3). Here, $\tilde{\mathbf{P}}_k$ and $\tilde{\mu}$ are sparsified versions of their dense counterparts. The regularisation parameter $\eta \geq 0$ controls the deviation of $\mathbf{c}$ from zero to penalise the reconstruction of highly unlikely dense displacement fields far away from μ. This is a common way to reduce the effects of noise and sparsity on dense motion reconstruction (see [24], [25] as examples).

C. Coupled Convex Optimisation for Improved Correspondence Estimation

To jointly minimise the block-matching dissimilarity and the model-based regularisation cost, we have adopted an alternating global optimisation method [36], [37]. By introducing an auxiliary vector $\tilde{\mathbf{v}}_t$, the dissimilarity term $\mathcal{D}$ and regularisation term $\mathcal{R}$ of (1) can be decoupled and an extra coupling term can be added, so that minimisation of this new cost function is equivalent to the minimisation of (1) for $\theta \to \infty$:

$$E(\tilde{\mathbf{u}}_t, \tilde{\mathbf{v}}_t) = \sum_{\Omega_N} \mathcal{D}(I_R, I_M, \tilde{\mathbf{u}}_t) + \theta \|\tilde{\mathbf{u}}_t - \tilde{\mathbf{v}}_t\|_2^2 + \alpha \mathcal{R}(\tilde{\mathbf{v}}_t) \ . \quad (5)$$

The cost function in (5) is initialised with $\theta = 0$ at first and for the regularisation term $\mathcal{R}$, we implicitly select α by determining the number of eigenvectors k for the model. The initial sparse motion field $\tilde{\mathbf{u}}_t$ is computed using the block-matching algorithm under unconstrained conditions, i.e. $\theta = 0$. The regularised sparse field $\tilde{\mathbf{v}}_t$ can then be computed by projecting $\tilde{\mathbf{u}}_t$ into model space by estimating weight vector $\mathbf{c}$ using (4) and computing $\tilde{\mathbf{v}}_t$ with the sparse model $\tilde{\mathbf{P}}_k \mathbf{\Sigma}_k$ and $\mathbf{c}$ as in (3). Now the stored dissimilarity map from the block-matching result is updated by adding the coupling term $\theta \|\tilde{\mathbf{u}}_t - \tilde{\mathbf{v}}_t\|_2^2$ and using this combined cost we can determine the new sparse motion field $\tilde{\mathbf{u}}_t$. This alternating process is iterated to find an optimal sparse displacement vector field, which minimises the entire cost. In practice, the cost function quickly converges after few iterations by employing a gradually increasing parameter θ.

In Fig. 2, an example of sparse displacement field after both optimisation steps is shown. Although the regularised displacement vectors after the first iteration (b) are more consistent than those after block-matching (a), some displacement vectors are still erroneous. With each optimisation step, the displacement vectors become more spatially consistent (c) as expected for physiological motion (d).

In addition to the spatial regularisation, we can easily incorporate prior knowledge about temporally smooth motion into the optimisation. This is done by extending (5) with an additional term $\beta \|\tilde{\mathbf{u}}_t - \tilde{\mathbf{u}}_{t-1}\|_2^2$, where we penalise differences between the displacements estimated at time point t and $t-1$:

$$E = \sum_{\Omega_N} \mathcal{D} + \theta \|\tilde{\mathbf{u}}_t - \tilde{\mathbf{v}}_t\|_2^2 + \beta \|\tilde{\mathbf{u}}_t - \tilde{\mathbf{u}}_{t-1}\|_2^2 + \alpha \mathcal{R} \ . \quad (6)$$

The estimated motion is then also temporarily smooth.

III. Experiments and Results

For the evaluation of our method and comparisons to state-of-the-art, we have used the following three publicly available data sets, complemented by us with a substantial amount of manually set landmarks[1].

[1]The landmarks were manually selected and verified by the authors of this work, who all have several years of experience in thoracic/abdominal image analysis. The different numbers of landmarks used for each data set are mainly a result of the different image qualities and field of views (FOV). Broadly speaking, we tried to identify as many clearly trackable landmarks as possible in every image sequence. However, especially for the US data it is very difficult to identify more than 1 or 2 distinguishable landmarks because of the SNR and the limited FOV (see also [18]).

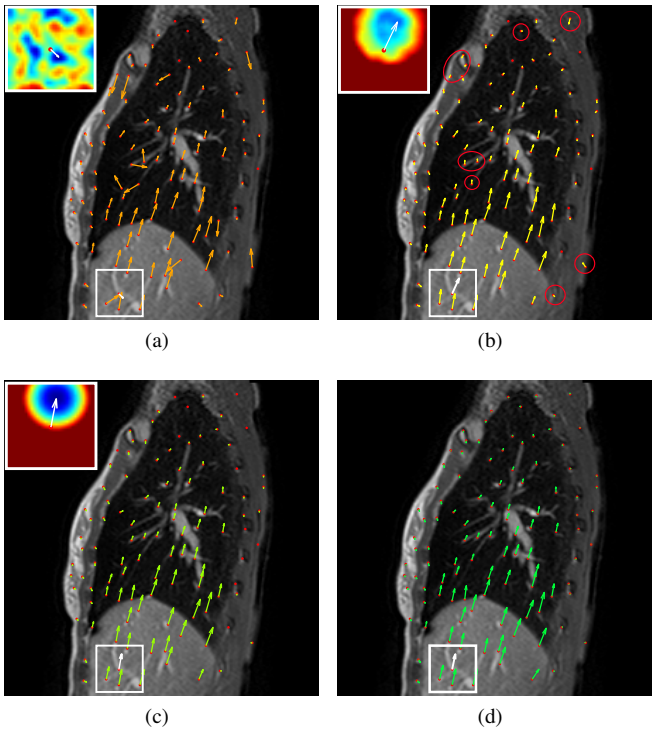

Fig. 2. Examples of the sparse displacement field after block-matching (BM) (a), first (b) and sixth iteration (c) of coupled convex optimisation (CCO) and the sparse displacement fields illustrating the physiological motion based on the *deeds* registration (d) are shown. On the left top of each image, the dissimilarity map with the minimal cost pointed by the arrow is also depicted, where the warm color indicate high cost and the cold color low cost. The displacement vectors estimated by block-matching contains erroneous displacement vectors, which are not consistent with the global motion. Some errors are corrected after the regularisation (first iteration), however, still some displacement vectors do not match the global motion (in red circles). After sixth iteration, the sparse motion field is almost identical to the *deeds* registration result.

a) 2D+t MRI: Four 2D+t T1-weighted MRI data sets (*volA*–*volD*) of the thorax/abdomen of 4 different patients from [33]. The data available for each patient consist of 30 2D+t images series (40 time points) acquired at different sagittal slice positions during free breathing (see [33] for details). Each 2D sagittal image has 215×288 voxels with an isotropic voxel spacing of 1.39 mm. For the evaluation, the 2D+t data for 2 sagittal slice positions (left lung/right lung center) are selected for each patient (= 8 2D+t images series in total). 10–20 corresponding landmarks in the lung/liver region were manually determined in 5 randomly selected 2D images of each 2D+t image series. The landmarks and MATLAB code is available under https://github.com/mattiaspaul/realtimeDeeds.

b) 4D MRI: Two 4D MRI data sets (*sl010* and *sl014*) of the thorax/abdomen from [23]. Each data set consist of 200 frames of 3D volume images taken from a healthy volunteer under free breathing for several respiratory circles. The 3D images have an isotropic inplane (sagittal) spatial resolution of 224×224 voxels with a spacing of 1.21 mm/1.30 mm and an inter-slice distance of 5 mm for 50 and 52 slices, respectively. For each data set, 27 corresponding landmarks were manually determined in the lungs and the liver in 10–11 randomly selected image frames. For this data, only sagittal inplane

motion is estimated due to the large inter-slice distance.

c) 4D US: Nine 4D US data sets (*SMT01–09*) of liver from the CLUST Challenge [18] containing 92–96 3D image frames with an isotropic spatial resolution of 0.7 mm (image size: $227\times229\times229$ voxels) and a temporal resolution of 8 Hz. For the evaluation, we have used manually set 1–2 landmarks in 9–10 randomly selected frames of each data set.

We first evaluate each component of our method with different parameter values to determine the effect of them using a subset of the 4D MRI data *b)*. Using these optimal parameters, we then perform a thorough evaluation on all MRI data (*a)* and *b)*) and compare the results with a state-of-the-art method of [30]. Finally, we evaluate our method for the 4D US data *c)*. All comparisons are based on mean target registration errors (TRE) calculated by using the manually determined landmarks.

In all experiments, the first frame of each data set is chosen as the reference image, to which the remaining image frames were registered using the *deeds* registration algorithm [35] to obtain dense displacement fields and for which landmarks are available. To assess the statistical significance of differences between overall mean TRE of different methods, we performed paired t-tests with a significance level of 5% ($p < 0.05$) by paring the patient-specific mean TRE.

A. Effect of different parameters

A subset of the 4D MRI data (*sl010*) is used in this experiment, where displacement fields of the first third of the images are used to estimate a patient-specific motion model (*training data*) and the remaining images were used to represent the intraoperative image acquisition phase, i.e. the online images (*test data*).

a) Feature points and block-matching: For evaluating the accuracy of sparse-to-dense extrapolation of our method, different configurations of the Harris corner detector are compared. We have varied the radius for non-maximum suppression, which affects the number of feature points selected by the Harris detector. The feature points are extracted from three selected slices, since the first clinically available MRI-guided radiotherapy system (Viewray) can acquire 3 parallel frames simultaneously at 2 fps rate [38].

The block-size for our block-matching algorithm is set to 11 pixels and for the coupled convex optimisation parameter θ, logarithmically increasing values between 0.03 to 1 are selected. The size of the search regions of each data set for the block-matching is determined based on the maximal displacement observed in automatic registration of the pre-treatment phase. The dependency of our motion estimation on the number of feature points are shown in Fig. 3. Although the reconstruction error tends to decreases after the block-matching as the number of feature points increases, together with the proposed coupled convex optimisation method, it only has small influence on the motion estimation accuracy. The optimal number of feature points per image slice for the data set is around 100 points.

b) Coupled convex optimisation: In Fig. 4, mean estimation errors (blue line) for the test data using different numbers of iterations for coupled convex optimisation are shown along with the accumulated computation time (dotted line).

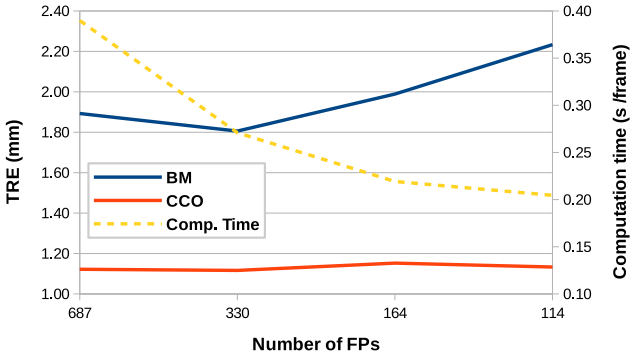

Fig. 3. Mean landmark errors of the estimated dense motion fields reconstructed using different numbers of feature points. The TRE is computed for 27 landmarks from each of 10 selected frames. All feature points are extracted from three MRI slices and the total computation time for block-matching and 6 iteration of coupled convex optimisation is given. When employing our proposed coupled convex optimisation method, the influence of the sparsity on the TRE is small and fewer feature points are required.

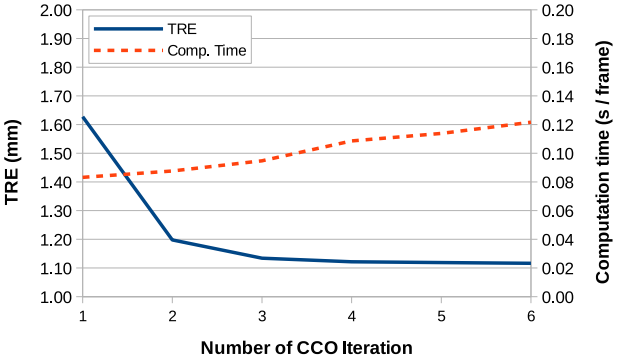

Fig. 4. Mean landmark estimation error using coupled convex optimisation with different number of iteration. TRE is computed for 27 landmarks from each of 10 selected frames and 330 feature points are tracked using block-matching. The blue line indicate the mean error of landmarks and the red dotted line shows the computation time of coupled convex optimisation.

When using the model-based regularisation only once, our method would be comparable to [24]. However, we further optimise the cost function based on (5) using the coupled convex optimisation. While this clearly improves the registration accuracy the computation time is only marginally increased for each additional iteration (amounting to about 40 ms for six iterations).

B. Evaluation on MRI Data

Next, we use both 4D MRI data sets, *sl010* and *sl014* and all four 2D+t MRI data sets (*a)* and *b)*) for the quantitative evaluation of our method and a comparison to the current state-of-the-art method for optical flow based registration called *RealTITracker* [30]. It can register 2D, 2D multislice, and 3D images and was developed for MRI-based realtime tracking in a MR-Linac. An extension employing a PCA model is also available, which ensures the spatio-temporal coherency

of the periodic organ motion in the registration through a learning step [27]. For the comparison, we have chosen the *RealTITracker* for 2D multislice registration with the PCA extension and used a PCA basis of size 4 as recommended in [27].

For our method the settings described in Section III-A are used. For the 2D+t data set, a temporally smooth motion is estimated by adding the second regularisation term $\beta||\tilde{\mathbf{u}}_t - \tilde{\mathbf{u}}_{t-1}||_2^2$ (6). This term penalises temporal deviation from the previous frame and the motion estimation result can be improved with appropriate parameters. The parameter β is set to 0.5 for the first iteration, 0.25 for the second iteration and 0 for the rest iterations. For the training of the 2D+t MRI data sets, we have used half of the images (20) and the remaining images (20) were used as the test data.

In Table I the TRE of the landmarks are shown for 4D and 2D+t MRI data sets. The t-test results of for 4D MRI results show that the TRE of our method only have insignificant differences compared to that of the *deeds* registration and are significantly better than the TRE of the block-matching (both $\eta = 0$ and $\eta > 0$). For 4D MRI, the results are also significantly better than the compared state-of-the-art algorithm (*RealTITracker*) and the sliding motion of the organs can be much better reproduced using our method as shown in Fig. 5.

In addition to calculating the mean TRE reported in Table I, we also quantitatively evaluated the ability to correctly handle sliding motion of all tracking approaches. This was done by specifically analysing the TRE for the additional landmarks located at bony structures (mainly ribs and vertebrae), which either do not move at all (vertebrae) or whose motion is mainly orthogonal to the predominant cranio-caudal motion of the lungs and the liver (ribs). The mean TRE for these landmarks in our MRI images is 3.34 mm using *RealTITracker* and 1.65 mm using our proposed method. This result clearly shows the advantage of our approach over the RealTITracker in sliding motion scenarios.

Another important aspect is the estimation of motion in mid-ventilation phases, which is often more challenging than the extreme phases such as in-/exhale phases, since it contains more motion artefacts [39]. Using the *RealTITracker*, the TRE of mid-ventilation phases (2.24 mm) tends to have higher values than the in-/exhale phases (1.57 mm). However, the TRE of *deeds* registration as well as our proposed method have similar values for both mid-ventilation and extreme phases (*deeds*: 1.11 mm vs. 1.00 mm; our method: 1.13 mm vs. 1.04 mm respectively).

For the results of 2D+t MRI data sets, our method again shows a significant improvement in accuracy compared to the block-matching (both $\eta = 0$ and $\eta > 0$) and a similar motion compensation rate[2] ($\approx 84\%$) as the *deeds* registration ($\approx 87\%$).

Moreover, our method can be computed in realtime, approximately 300 times faster ($\approx$ 0.2 s for 4D MRI) than the *deeds* registration on a CPU, which takes more than a minute to process an image.

[2]motion compensation rate = (A-B)/A*100
A = mean Euclidean distance of landmark x in I_R and I_T before motion compensation
B = mean Euclidean distance of x in I_R and I_T after motion compensation

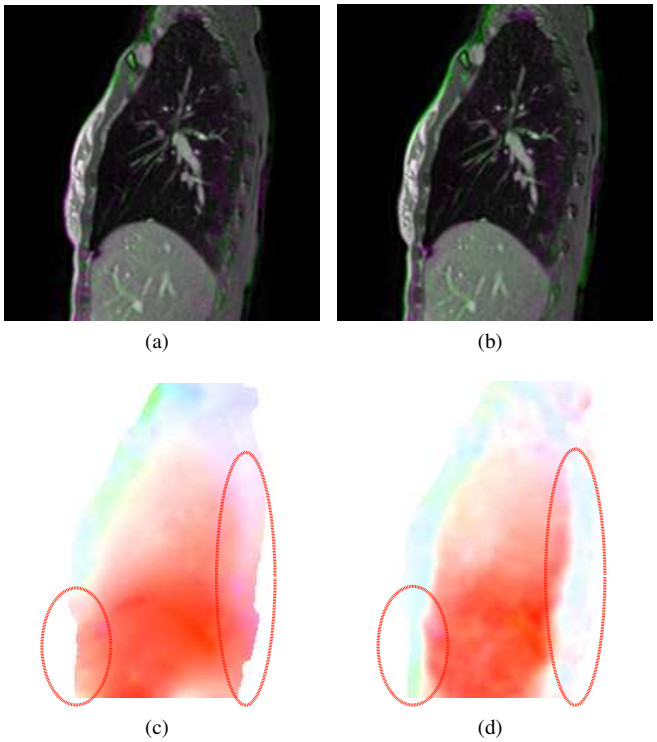

Fig. 5. Motion estimation for sliding motion. The difference image between reference image and registered image using the *RealTITracker* (a) and our method (b) are shown. The optical flow illustrations for corresponding slice are shown on the bottom ((c) and (d)). The true motion of regions marked with a red circle (including skin and bones) is closed to 0 in cranio-caudal direction during respiration. Our method (d) correctly predicts this sliding motion.

C. Evaluation on Ultrasound Data

In addition to the MRI data sets, we also evaluated our method on 4D US data sets (*SMT01–09*). The tracking of feature points in US images is usually more challenging than in MRI images, as some structures might disappear from the image field-of-view, due to for example, the respiratory motion or the movement of the ultrasound probe. Moreover, the image quality of the US is much lower than that of the MRI images.

Because the frame rate of the ultrasound images is much higher than that of an MRI scanner, we have added the temporal regularisation of the motion for this experiment as described in (6). The block-size was set to 7 pixels and the coupled convex optimisation scheme is performed with logarithmically increasing θ values between 0.02 – 5.0. For the temporal regularisation we have set the same β values as in the experiment of 2D+t MRI.

As shown in Table II, both of our proposed methods (with and without temporal regularisation) significantly improve the accuracy compared to the block-matching algorithm ($\eta = 0$ and $\eta > 0$). The achieved motion compensation rate of our method ($\approx 77\%$) is also comparable to the *deeds* registration ($\approx 78\%$). However, the adaptation of temporal regularisation does not improve the accuracy of our method significantly.

TABLE I
EXPERIMENTAL RESULTS OF 2D+T MRI AND 4D MRI DATA SETS: TRE (MEAN LANDMARK DISTANCE IN MM) AND COMPUTATION TIME (SECOND/FRAME) OF *deeds* REGISTRATION, BLOCK-MATCHING (BM), REGULARISATION AFTER BLOCK-MATCHING ($\eta > 0$) AND OUR PROPOSED METHOD WITH AND WITHOUT TEMPORAL COMPONENT.

Data set	without registration	*deeds* registration	BM ($\eta = 0$)	BM ($\eta > 0$)	Model-based w/o temporal	Model-based w/ temporal	*RealTITracker 2D multi w/ PCA*
			4D MRI				
sl010	3.97 ± 2.33	1.12 ± 0.62	1.81 ± 1.16	1.63 ± 1.02	1.12 ± 0.64	–	2.07 ± 1.29
sl014	1.78 ± 1.04	1.01 ± 0.52	1.54 ± 0.94	1.26 ± 0.66	1.07 ± 0.57	–	1.67 ± 0.92
mean TRE (mm)	2.87	**1.06**	1.72	1.45	**1.10**	–	**1.87**
CPU comp. time (s/frame)	–	$\approx$ **60.0**	≈ 0.100	≈ 0.127	$\approx$ **0.204**	–	$\approx$ **3.82**
GPU comp. time (s/frame)		–	≈ 0.027	≈ 0.036	$\approx$ **0.034**		–
			2D+t MRI				
Volunteer A	6.26 ± 3.47	0.97 ± 1.00	1.77 ± 0.99	1.60 ± 1.41	1.13 ± 1.27	1.27 ± 0.99	1.12 ± 0.88
Volunteer B	3.47 ± 1.85	0.59 ± 0.31	0.97 ± 0.33	0.92 ± 0.44	0.73 ± 0.46	0.72 ± 0.42	0.82 ± 0.37
Volunteer C	8.09 ± 2.55	0.73 ± 0.40	1.32 ± 0.44	1.32 ± 0.68	0.87 ± 0.72	0.85 ± 0.47	0.83 ± 0.74
Volunteer D	5.65 ± 3.65	0.71 ± 0.69	1.75 ± 0.68	1.65 ± 1.23	1.01 ± 1.11	0.96 ± 0.76	1.37 ± 1.19
mean TRE (mm)	5.87	**0.75**	1.45	1.37	0.94	**0.95**	**1.04**
CPU comp. time (s/frame)	–	$\approx$ **9.15**	≈ 0.100	≈ 0.101	≈ 0.114	$\approx$ **0.118**	$\approx$ **0.31**
GPU comp. time (s/frame)		–	≈ 0.001	≈ 0.006	≈ 0.003	$\approx$ **0.004**	–

Our resulting TRE of 2.21 mm is close to the average inter-observer distance, which we measured to be 1.67 mm based on a second rater.

Although our method also improves the computation time in this experiment, it is substantially higher compared to the MRI experiments, because dissimilarities are computed for a dense 3D displacement space, requiring more than half a second for motion estimation of an image frame. To achieve realtime performance, the computation time should be at least under 125 ms (8 Hz), which can be realised by a GPU implementation of our method as explained next.

D. GPU Implementation

The full reconstruction of a displacement field using our algorithm takes up to 5 seconds for an image on a CPU using standard block-matching algorithm (Intel Core i5-6600 @ 3.30 GHz (4 cores) with 32 GB RAM). Block-matching and coupled convex optimisation are the most time consuming components of our framework. Fortunately, we can reduce the computation time of the block-matching algorithm using convolution filters, which enables a straightforward parallelisation of the computation of local image similarities and avoids sequential calculations.

The computation of the sum of squared differences between two SSC descriptor patches P and Q can be broken down into the addition of three terms as follows:

$$\frac{1}{\Omega_p} \sum_{i,j \in \Omega_p} (P_{ij} - Q_{ij})^2 = \frac{1}{\Omega_p} \sum_{i,j \in \Omega_p} (P_{ij}^2 - 2P_{ij}Q_{ij} + Q_{ij}^2), \tag{7}$$

where Ω_p denotes the size of the image patch, i.e. the block-size and P_{ij} and Q_{ij} denote the SSC descriptor values at (i, j). For the patches in the search region, the correlation between the two patches (P and Q) can be computed at once for all feature points via convolution, where the reference image patches are considered as groups of filters. It reduces the computation time approximately 6 times when implemented on a CPU and with the GPU implementation of the same algorithm, it achieves a realtime performance and total speed-up of more than one order of magnitude as shown in Table I and Table II (experiments were performed using an NVIDIA TITAN XP[3]).

In addition, the motion field reconstruction, coupled convex optimisation as well as the SSC descriptor computation can be straightforwardly implemented on the GPU, further reducing the computation time to $<$ 50 ms for both MRI data sets and $<$ 120 ms for the US data set.

IV. DISCUSSION

In this work, we have presented a novel approach to registration-based realtime estimation of patient motion in image-guided radiotherapy without any markers. By introducing a coupled-convex optimisation our method is able to integrally combine efficient and flexible block-matching dissimilarity distributions for sparsely sampled feature points and patient-specific motion models.

Given an appropriate highly-accurate dense registration method (in this case *deeds* [35]), we can obtain high-quality motion models of a pre-treatment phase that can deal with sliding motion and accurately estimate both tumour and organ-at-risk motion.

The extensive quantitative results on three separate publicly available datasets that represent use-cases for IGRT in both MRI and ultrasound, demonstrate that our sparsified jointly optimised model nearly reaches the lower-bound of accuracy compared to the registration quality of the dense approach, while reducing computations by up to 100×. This speed-up could be obtained through the combination of sparse sampling and an elegant GPU implementation of the dissimilarity maps using convolution filters.

The detailed experiments clearly show that a disconnected post regularisation of block-matching vectors as proposed in previous work is not sufficient and can be significantly out-

[3]We gratefully acknowledge the support of NVIDIA Corporation with the donation of the Titan X Pascal GPU used for this research.

TABLE II
EXPERIMENTAL RESULT OF 4D US DATA SETS: TRE (MEAN LANDMARK DISTANCE IN MM) AND COMPUTATION TIME (SECOND/FRAME) OF *deeds* REGISTRATION, BLOCK-MATCHING (BM), REGULARISATION AFTER BLOCK-MATCHING ($\eta > 0$) AND OUR PROPOSED METHOD WITH AND WITHOUT TEMPORAL COMPONENT.

Data set (landmark)	without registration	*deeds* registration	BM ($\eta = 0$)	BM ($\eta > 0$)	Model-based w/o temporal	Model-based w/ temporal
SMT01 (1)	7.86 ± 3.38	0.93 ± 0.18	2.63 ± 1.25	1.60 ± 1.21	1.04 ± 0.33	0.99 ± 0.30
SMT01 (2)	8.31 ± 2.45	1.35 ± 0.73	2.56 ± 0.89	1.92 ± 0.65	1.44 ± 0.86	1.43 ± 0.85
SMT02 (1)	6.21 ± 2.63	0.88 ± 0.09	2.65 ± 1.67	1.05 ± 0.38	0.90 ± 0.19	0.90 ± 0.19
SMT03 (1)	6.07 ± 4.33	0.72 ± 0.22	3.76 ± 1.87	1.95 ± 2.11	1.20 ± 1.10	1.04 ± 0.68
SMT04 (1)	14.20 ± 7.44	6.18 ± 5.51	9.91 ± 5.26	6.76 ± 5.44	6.52 ± 5.46	6.63 ± 5.35
SMT05 (1)	12.30 ± 7.01	3.05 ± 1.42	6.23 ± 1.94	5.46 ± 3.32	4.30 ± 2.87	3.24 ± 1.40
SMT06 (1)	15.91 ± 5.66	2.07 ± 0.73	6.50 ± 4.21	4.40 ± 3.29	3.14 ± 2.40	2.35 ± 1.00
SMT06 (2)	14.16 ± 7.99	4.64 ± 3.04	8.84 ± 3.69	7.59 ± 6.26	6.44 ± 5.72	4.73 ± 3.72
SMT07 (1)	5.35 ± 2.20	1.12 ± 0.52	1.85 ± 0.71	1.35 ± 0.58	1.19 ± 0.43	1.19 ± 0.47
SMT08 (1)	4.20 ± 2.63	0.63 ± 0.21	1.84 ± 0.78	0.70 ± 0.32	0.63 ± 0.19	0.62 ± 0.20
SMT09 (1)	9.46 ± 3.01	1.08 ± 0.65	3.10 ± 1.65	1.51 ± 0.99	1.19 ± 0.76	1.14 ± 0.69
mean	9.46	**2.06**	4.53	3.12	**2.54**	**2.20**
comp. time CPU (s/frame)	–	$\approx$ **60.0**	≈ 0.160	≈ 0.265	≈ 0.612	$\approx$ **0.610**
comp. time GPU (s/frame)		–	≈ 0.020	≈ 0.042	≈ 0.109	$\approx$ **0.116**

performed by our proposed incorporation of discrete matching and model-based regularisation and its joint optimisation.

Compared to our initial work presented as conference paper [32], we performed a detailed analysis of the effects of the different parameters such as the number of feature points to be tracked (sparsity of the online information) and the number of iterations for the proposed coupled convex optimisation (CCO). It was shown that influence of the number of feature points and hence the sparsity of the model is greatly reduced when including the regularisation with CCO. The convergence of this alternating optimisation was also shown both visually and quantitatively.

Our results for multi/slice MRI data compare favourably for both accuracy and computation time with the the advanced optical flow technique [30] (together with PCA motion models [27]) and we have therefore set a new state-of-the-art in realtime MRI tracking.

The proposed method will in particular have a great impact on IGRT by providing a viable solution to the limited acquisition time in MRI-Linac systems based on our sparse-to-dense motion field reconstruction.

Due to the flexibility of our approach a general discussion of our results in terms of their specific clinical impact is challenging. However, it is safe to say that, for example, in radiation therapy, perfect target tracking (tracking error of 0) would render safety margins obsolete. Hence, aside from technical restrictions, any improvement in tracking accuracy will directly lead to smaller safety margins and will consequently improve (and possibly speed-up) dose delivery to the tumour while simultaneously sparing healthy surrounding tissue. According to the recommendations of AAPM Task Group 76 for guided radiation therapy treatments [2], the total time delay of a real-time tracking/compensation systems should be kept as short as possible and in no case exceed 0.5 seconds. We achieved computation times of < 50 ms (MRI-based tracking) and < 120 ms (US-based tracking) substantially below the required threshold and therefore the rest of the available processing time can be used for other parts of the whole pipeline (image acquisition, collimator adaption, ...).

One limitation of our work is that our experiment on 4D MRI dataset only accounts for inplane motion. This is due to the large inter-slice spacing of the data used. However, if MRI images with smaller inter-slice spacing is available in the treatment planning phase, we believe that it is more reasonable to track 3D motion and our method can be easily applied for 3D motion tracking as shown in the experiment on 4D US dataset. Another limitation of our current implementation is the assumption that the respiratory motion pattern of the patient does not change drastically between the pre-treatment phase and the actual treatment and therefore the search region used for block-matching can be restricted accordingly. However, this may not always be guaranteed in clinical practice due to inter- and intra-fractional motion variability. In future work, we will, therefore, investigate the integration of online model and block-matching adaptation strategies to deal with problems like baseline shifts [40]. Additionally, it might also be possible to integrate temporal information into model building (e.g., spatio-temporal PCA), which could in future even replace our temporal constraint.

V. CONCLUSIONS

Our work presents a novel solution to dense patient motion estimation for non-invasive IGRT in realtime. We demonstrate that by jointly optimising a discrete GPU-accelerated block-matching step for sparsely distributed feature points combined with a patient-specific motion-model for regularisation, a very accurate sparse-to-dense motion estimation covering both tumour and organs-at-risk can be accurately computed in few milliseconds ($100\times$ speed-up compared to naive implementation). The individual algorithmic contributions are extensively validated using three different challenging temporal MRI and ultrasound datasets of lungs and liver, which are important areas for radiation therapy that require accurate motion compensation. We achieve significant improvements in computation speed and accuracy compared to disjoint optimisation within the same (block-matching) framework and also outperform the

current state-of-the-art optical-flow based registration methods [30].

REFERENCES

[1] J. Blackall et al., "MRI-based measurements of respiratory motion variability and assessment of imaging strategies for radiotherapy planning," Physics in Medicine and Biology, vol. 51, no. 17, p. 4147, 2006.

[2] P. J. Keall et al., "The management of respiratory motion in radiation oncology report of AAPM Task Group 76," Med Phys, vol. 33, no. 10, pp. 3874–3900, 2006.

[3] Y. Zhang et al., "Deformable motion reconstruction for scanned proton beam therapy using on-line X-ray imaging," Physics in Medicine and Biology, vol. 58, no. 24, p. 8621, 2013.

[4] M. Wilms et al., "Multivariate regression approaches for surrogate-based diffeomorphic estimation of respiratory motion in radiation therapy," Phys Med Biol, vol. 59, no. 5, pp. 1147–64, 2014.

[5] B. D. de Senneville et al., MRI-Guided High-Intensity Focused Ultrasound Sonication of Liver and Kidney. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 349–366.

[6] D. Fontanarosa et al., "Review of ultrasound image guidance in external beam radiotherapy: I. Treatment planning and inter-fraction motion management," Phys Med Biol, vol. 60, no. 3, pp. R77–R114, 2015.

[7] J. J. W. Lagendijk et al., "MR guidance in radiotherapy," Phys Med Biol, vol. 59, no. 21, pp. R349–R369, 2014.

[8] J. Ruhaak et al., "Estimation of large motion in lung CT by integrating regularized keypoint correspondences into dense deformable registration," IEEE Transactions on Medical Imaging, vol. 36, pp. 1746–1757, 2017.

[9] M. P. Heinrich et al., "Towards realtime multimodal fusion for image-guided interventions using self-similarities," in International conference on medical image computing and computer-assisted intervention. Springer, 2013, pp. 187–194.

[10] B. Stemkens et al., "Image-driven, model-based 3D abdominal motion estimation for MR-guided radiotherapy," Physics in Medicine and Biology, vol. 61, no. 14, p. 5335, 2016.

[11] T. Bjerre et al., "Three-dimensional MRI-linac intra-fraction guidance using multiple orthogonal cine-MRI planes," Physics in Medicine and Biology, vol. 58, no. 14, p. 4943, 2013.

[12] L. I. Cerviño et al., "MRI-guided tumor tracking in lung cancer radiotherapy," Physics in Medicine and Biology, vol. 56, no. 13, p. 3773, 2011.

[13] V. De Luca et al., "Speeding-up image registration for repetitive motion scenarios," in IEEE International Symposium on Biomedical Imaging 2012. IEEE, 2012, pp. 1355–1358.

[14] L. Brix et al., "Three-dimensional liver motion tracking using real-time two-dimensional MRI," Medical Physics, vol. 41, no. 4, pp. 042 302–n/a, 2014, 042302.

[15] C. Paganelli et al., "Magnetic resonance imagingguided versus surrogate-based motion tracking in liver radiation therapy: A prospective comparative study," International Journal of Radiation Oncology*Biology*Physics, vol. 91, no. 4, pp. 840 – 848, 2015.

[16] V. De Luca et al., "A learning-based approach for fast and robust vessel tracking in long ultrasound sequences," in MICCAI 2013. Springer Berlin Heidelberg, 2013, vol. 8149, p. 518.

[17] J. Banerjee et al., "Fast and robust 3D ultrasound registration block and game theoretic matching," Medical Image Analysis, vol. 20, no. 1, pp. 173 – 183, 2015.

[18] V. D. Luca et al., "The 2014 liver ultrasound tracking benchmark," Physics in Medicine and Biology, vol. 60, no. 14, p. 5571, 2015.

[19] A. J. Shepard et al., "A block matching based approach with multiple simultaneous templates for the real-time 2D ultrasound tracking of liver vessels," Medical Physics, pp. n/a–n/a, 2017, in press.

[20] D. Lee and A. Krupa, "Intensity-based visual servoing for non-rigid motion compensation of soft tissue structures due to physiological motion using 4D ultrasound," in 2011 IEEE/RSJ International Conference on Intelligent Robots and Systems, Sept 2011, pp. 2831–2836.

[21] L. Royer et al., "Real-time target tracking of soft tissues in 3D ultrasound images based on robust visual information and mechanical simulation," Medical Image Analysis, vol. 35, pp. 582 – 598, 2017.

[22] J. McClelland et al., "Respiratory motion models: A review," Medical Image Analysis, vol. 17, no. 1, pp. 19 – 42, 2013.

[23] D. Boye et al., "Population based modeling of respiratory lung motion and prediction from partial information," in SPIE medical imaging. International Society for Optics and Photonics, 2013, pp. 86 690U–86 690U.

[24] F. Preiswerk et al., "Model-guided respiratory organ motion prediction of the liver from 2D ultrasound," Medical Image Analysis, vol. 18, no. 5, pp. 740 – 751, 2014.

[25] T. Klinder and C. Lorenz, "Respiratory motion compensation for image-guided bronchoscopy using a general motion model," in Biomedical Imaging (ISBI), 2012 9th IEEE International Symposium on. IEEE, 2012, pp. 960–963.

[26] A. King et al., "Thoracic respiratory motion estimation from MRI using a statistical model and a 2-D image navigator," Medical Image Analysis, vol. 16, no. 1, pp. 252 – 264, 2012.

[27] B. D. de Senneville et al., "A direct PCA-based approach for real-time description of physiological organ deformations," IEEE Transactions on Medical Imaging, vol. 34, no. 4, pp. 974–982, April 2015.

[28] O. Somphone et al., "Live feature tracking in ultrasound liver sequences with sparse demons," in MICCAI Challenge on Liver Ultrasound Tracking CLUST 2014, 2014, pp. 53–60.

[29] M. Seregni et al., "A hybrid image registration and matching framework for real-time motion tracking in MRI-guided radiotherapy," IEEE Transactions on Biomedical Engineering, 2017, in press.

[30] C. Zachiu et al., "An improved optical flow tracking technique for real-time MR-guided beam therapies in moving organs," Physics in medicine and biology, vol. 60, no. 23, p. 9003, 2015.

[31] L. König et al., "A non-linear image registration scheme for real-time liver ultrasound tracking using normalized gradient fields," in MICCAI Challenge on Liver Ultrasound Tracking CLUST 2014, 2014, p. 29.

[32] M. Wilms et al., "Model-based regularisation for respiratory motion estimation with sparse features in image-guided interventions."

[33] C. F. Baumgartner et al., "Autoadaptive motion modelling for MR-based respiratory motion estimation," Medical Image Analysis, vol. 35, pp. 83 – 100, 2017.

[34] C. Harris and M. Stephens, "A combined corner and edge detector." in Alvey vision conference, vol. 15, no. 50. Citeseer, 1988, pp. 10–5244.

[35] M. P. Heinrich et al., "MRF-based deformable registration and ventilation estimation of lung CT," IEEE Transactions on Medical Imaging, vol. 32, no. 7, pp. 1239–1248, July 2013.

[36] F. Steinbrücker et al., "Large displacement optical flow computation withoutwarping," in Computer Vision, 2009 IEEE 12th International Conference on. IEEE, 2009, pp. 1609–1614.

[37] M. P. Heinrich et al., "Non-parametric discrete registration with convex optimisation," in International Workshop on Biomedical Image Registration. Springer International Publishing, 2014, pp. 51–61.

[38] S. Mutic, "WE-A-BRA-02: First Commercial Hybrid MRI-IMRT System," Medical Physics, vol. 39, p. 3934, 2012.

[39] T. N. van de Lindt et al., "An MRI-based mid-ventilation approach for radiotherapy of the liver," Radiotherapy and Oncology, vol. 121, no. 2, pp. 276–280, 2016.

[40] D. Ruan et al., "Real-time profiling of respiratory motion: baseline drift, frequency variation and fundamental pattern change," Physics in medicine and biology, vol. 54, no. 15, p. 4777, 2009.