Real-Time Respiratory Motion Analysis Using Manifold Ray Casting of Volumetrically Fused Multi-view Range Imaging

Jakob Wasza[1], Sebastian Bauer[1], and Joachim Hornegger[1,2]

[1] Pattern Recognition Lab, Department of Computer Science
[2] Erlangen Graduate School in Advanced Optical Technologies (SAOT)
Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany
{jakob.wasza,sebastian.bauer,joachim.hornegger}@cs.fau.de

Abstract. A novel real-time multi-sensor framework for range imaging (RI) based respiratory motion analysis in image guided interventions such as fractionated radiation therapy is presented. We constitute our method based upon real-time constraints in clinical practice and an analytic analysis of RI based elastic body surface deformation fields. For the latter, we show that the underlying joint rigid and non-rigid registration problem is ill-conditioned and identify insufficient body coverage as an error source. Facing these issues, we propose a novel manifold ray casting technique enabling the reconstruction of an 180° coverage body surface model composed of $\sim 3 \cdot 10^5$ points from volumetrically fused multi-view range data in ~ 25 ms. Exploiting the wide field of view surface model enabled by our method, we reduce the error in motion compensated patient alignment by a factor of 2.7 in the translational and 2.4 in the rotational component compared to conventional single sensor surface coverage.

1 Introduction

Real-time respiratory motion analysis is a key issue for the success of medical procedures that require accurate patient positioning and respiratory motion prediction or compensation. For instance, in fractionated radiation therapy, the patient must be continuously monitored to account for tumour movement induced by respiratory motion [1]. For these tasks, multi-dimensional respiration surrogates based on elastic deformations of the patient body surface acquired by real-time range imaging (RI) sensors have been recently proposed [2–5]. These methods rely on fundamentally different surrogate generation concepts such as pure depth information [2], a model based formulation [5], non-rigid sparse to dense registration [3] or a non-rigid ICP variant [4]. Except for the work of Schaerer et al. [4] using the multi-sensor capable AlignRT system (Vision RT, London, UK), these approaches feature a single RI sensor setup. In this context, a noteworthy development is the recently proposed multi-sensor system by Price et al. [6] that, in contrast to the AlignRT system, features acquisition rates of > 20 fps thus allowing for high temporal resolution respiration analysis. However, existing multi-sensor systems (**i**) do not provide a unified fusion of data

K. Mori et al. (Eds.): MICCAI 2013, Part II, LNCS 8150, pp. 116–123, 2013.

streams from multiple cameras and (**ii**) are not capable to feature a fused surface representation that facilitates the deployment of high-performance algorithms.

In this paper, as the first and major contribution, we propose a novel real-time multi-sensor framework that enables a wide field of view patient surface model reconstruction. Our method employs a volumetric range data fusion approach encoding the patient body surface in an implicit manner. Based upon this representation we present an explicit surface reconstruction technique that utilizes a novel manifold ray casting technique tailored to the human anatomy present in respiration analysis. As a unique characteristic compared to existing approaches, our method enables both a high surface coverage and an efficient closest point search by exploiting inherent topological and projective properties. These features facilitate the usage of algorithms designed to cope with real-time constraints. Building upon this high coverage surface model, as the second contribution, we substantiate the necessity of a multi-sensor framework in RI based respiratory motion analysis. We show that the joint rigid and non-rigid registration problem that governs respiration analysis is ill-posed and identify insufficient body coverage as an error source. Consequently, as the third contribution, we demonstrate the efficiency of our framework regarding run-time and accuracy in the context of real-time respiration analysis using 4-D shape priors.

2 Method

We denote an RI as $f_{\mathcal{R}}(\boldsymbol{i}) : \Omega \rightarrow \mathbb{R}^+$ where the image domain $\Omega \subset \mathbb{R}^2$ is sampled with $N_1 \times N_2$ depth measurements $d_{\boldsymbol{i}} \in \mathbb{R}^+$. Without loss of generality, the RI can be equivalently described as a surface $f_{\mathcal{S}}(\boldsymbol{i}) : \Omega \rightarrow \Psi$ where the codomain $\Psi \subset \mathbb{R}^3$ holds the 3-D surface points $\boldsymbol{x}_{\boldsymbol{i}} \in \Psi$ as $\boldsymbol{x}_{\boldsymbol{i}} = \boldsymbol{o} + d_{\boldsymbol{i}} \cdot \boldsymbol{r}_{\boldsymbol{i}}$. The optical center $\boldsymbol{o}$ and the viewing rays $\boldsymbol{r}_{\boldsymbol{i}}$ are given by the pinhole camera model of the RI sensor. At respiration state k we denote the body surface $\mathcal{S}^k$ composed of N samples as a set of *fixed* points $\boldsymbol{x}_{\boldsymbol{i}}^{\mathcal{F}} \in \mathbb{R}^3$ from a reference respiration state (e.g. fully exhale) that are transformed by (1) an elastic deformation field $\mathcal{U}^k = \{\boldsymbol{u}_1^k, \ldots, \boldsymbol{u}_N^k\},\ \boldsymbol{u}_{\boldsymbol{i}}^k \in \mathbb{R}^3$ encoding the point-wise displacements induced by respiratory motion and (2) a rigid transformation $(\boldsymbol{R}^k \in \mathbb{SO}_3, \boldsymbol{t}^k \in \mathbb{R}^3)$ accounting for the patient pose:

$$\mathcal{S}^k = \{\boldsymbol{x}_1^k, \ldots, \boldsymbol{x}_N^k\} = \{\boldsymbol{R}^k\left(\boldsymbol{x}_1^{\mathcal{F}} + \boldsymbol{u}_1^k\right) + \boldsymbol{t}^k, \ldots, \boldsymbol{R}^k\left(\boldsymbol{x}_N^{\mathcal{F}} + \boldsymbol{u}_N^k\right) + \boldsymbol{t}^k\}. \quad (1)$$

Now, given an instantaneous RI surface $\mathcal{S}^{\mathcal{I}} = \{\boldsymbol{x}_1^{\mathcal{I}}, \ldots, \boldsymbol{x}_{N_1 \cdot N_2}^{\mathcal{I}}\},\ \boldsymbol{x}_i^{\mathcal{I}} \in \Psi$, the goal of RI based respiratory motion analysis techniques using elastic surface deformations is to recover the respiration state $\widehat{k}$ by jointly estimating the unknown displacement field $\mathcal{U}^{\widehat{k}}$ and the rigid transformation $(\boldsymbol{R}^{\widehat{k}}, \boldsymbol{t}^{\widehat{k}})$ such that $\mathcal{S}^{\mathcal{I}} \approx \mathcal{S}^{\widehat{k}}$ according to a certain shape distance criterion.

Let us note that the reference surface composed of *fixed* points $\boldsymbol{x}_{\boldsymbol{i}}^{\mathcal{F}}$ may be captured by RI sensors or computed from tomographic data (CT/MR) using isosurface extraction techniques. Thus, in general, RI based respiratory motion analysis is a multi-modal registration problem with $N \neq N_1 \cdot N_2$.

2.1 Volumetric Multi-View Range Data Fusion

For fusing multi-view RI data we use the concept of projective *truncated signed distance functions* (TSDF) [7, 8]. A TSDF $\mathcal{T}(\boldsymbol{p}) : \mathbb{R}^3 \rightarrow [-1, +1]$ is based on an implicit surface representation given by the zero level set of an approximated conventional signed distance function of that surface. For a point $\boldsymbol{p} \in \mathbb{R}^3$ in world space, $\mathcal{T}(\boldsymbol{p})$ encodes the distance to the closest point on the RI surface $\mathcal{S}$ w.r.t. the associated inherent projective camera geometry and the sensor pose:

$$\mathcal{T}(\boldsymbol{p}) = \eta\left(\|f_{\mathcal{S}}\left(\mathcal{P}\left(\boldsymbol{p}'\right)\right)\|_2 - \|\boldsymbol{p}'\|_2\right). \tag{2}$$

Here, $\boldsymbol{p}' = \boldsymbol{R}\boldsymbol{p} + \boldsymbol{t}$ with $(\boldsymbol{R} \in \mathbb{SO}_3, \boldsymbol{t} \in \mathbb{R}^3)$ is the transformation of $\boldsymbol{p}$ from world space into local camera space. $\mathcal{P}(\boldsymbol{p}') : \mathbb{R}^3 \rightarrow \Omega$ denotes the camera's projection operator and η performs mapping and truncation of distances to $[-1, +1]$. For a multi-view system with J sensors we extend Eq. (2) as:

$$\mathcal{T}(\boldsymbol{p}) = \frac{1}{\omega(\boldsymbol{p})} \cdot \sum_{j=1}^{J} f_{\mathcal{C}_j}\left(\mathcal{P}_j\left(\boldsymbol{p}'_j\right)\right) \cdot \eta\left(\left\|f_{\mathcal{S}_j}\left(\mathcal{P}_j\left(\boldsymbol{p}'_j\right)\right)\right\|_2 - \left\|\boldsymbol{p}'_j\right\|_2\right), \tag{3}$$

where $\omega(\boldsymbol{p}) = \sum_{j=1}^{J} f_{\mathcal{C}_j}\left(\mathcal{P}_j\left(\boldsymbol{p}'_j\right)\right)$. The *confidence* map $f_{\mathcal{C}_j}(\boldsymbol{i}) : \Omega_j \rightarrow \mathbb{R}^+$ associated with the j-th camera encodes the reliability of a surface measure $\boldsymbol{x_i} \in \Psi_j$. Given (1) the associated surface normal $\boldsymbol{n_i}$ and viewing ray $\boldsymbol{r_i}$, and (2) the gradient magnitude of the depth data $g_{\boldsymbol{i}} = \left\|\nabla f_{\mathcal{R}_j}(\boldsymbol{i})\right\|_2$, we propose:

$$f_{\mathcal{C}_j}(\boldsymbol{i}) = \langle \boldsymbol{n_i}, \boldsymbol{r_i} \rangle \cdot (1 + g_{\boldsymbol{i}})^{-1}. \tag{4}$$

This results in high scores for smooth surface regions that are perpendicular to the viewing direction. Finally, the temporal integration of TSDFs is given as:

$$\mathcal{T}(\boldsymbol{p})^{(k+1)} = \frac{(1-\alpha) \cdot \omega(\boldsymbol{p})^{(k-1)} \cdot \mathcal{T}(\boldsymbol{p})^{(k-1)} + \alpha \cdot \omega(\boldsymbol{p})^{(k)} \cdot \mathcal{T}(\boldsymbol{p})^{(k)}}{\omega(\boldsymbol{p})^{(k+1)}}, \tag{5}$$

where $\omega(\boldsymbol{p})^{(k+1)} = (1-\alpha) \cdot \omega(\boldsymbol{p})^{(k-1)} + \alpha \cdot \omega(\boldsymbol{p})^{(k)}$ and the parameter α controls the averaging of successive frames.

2.2 Surface Reconstruction Using Manifold Ray Casting

A TSDF $\mathcal{T}$ encodes the fused body surface in an implicit manner, however, the majority of surface processing techniques rely on an explicit representation, i.e. vertices and edges. For this task, surface reconstruction techniques based on ray casting the TSDF have proven to be real-time capable [8]. The basic idea is to define a virtual RI sensor and, starting at the optical center $\boldsymbol{o}$, to traverse $\mathcal{T}$ along the viewing rays $\boldsymbol{r_i}$ toward the zero level set of $\mathcal{T}$, i.e. $\mathcal{T}(\boldsymbol{o} + d_{\boldsymbol{i}} \cdot \boldsymbol{r}_i) = 0$. However, this conventional pinhole camera model does not allow for a wide field of view body coverage as depicted in Fig. 1(left). In contrast, by using a virtual camera model with viewing rays emerging from a 2-manifold $\mathcal{M}$, a high coverage

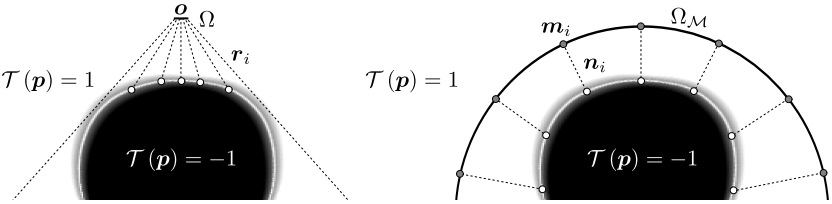

Fig. 1. Explicit surface generation by ray casting a TSDF $\mathcal{T}$. Pinhole camera model (left) and manifold ray casting based on a cylindrically shaped sensor plane (right).

body surface model can be reconstructed from the TSDF representation as illustrated in Fig. 1(right). The virtual RI domain $\Omega_\mathcal{M} \subset \mathbb{R}^2$ is given by a suitable 2-D parametrization of $\mathcal{M}$ which allows to define an RI as $f_{\mathcal{R}_\mathcal{M}}(\boldsymbol{i}) : \Omega_\mathcal{M} \to \mathbb{R}^+$ and the corresponding surface as $f_{\mathcal{S}_\mathcal{M}}(\boldsymbol{i}) : \Omega_\mathcal{M} \to \Psi_\mathcal{M}$. For an index $\boldsymbol{i} \in \Omega_\mathcal{M}$ that, using the 2-D parametrization, defines a point on the manifold $\boldsymbol{m_i} \in \mathcal{M}$ and the associated normal $\boldsymbol{n_i}$, a surface point $\boldsymbol{x_i} \in \Psi_\mathcal{M}$ is given as:

$$\boldsymbol{x_i} = \boldsymbol{m_i} + d_\boldsymbol{i} \cdot \boldsymbol{n_i}, \tag{6}$$

where the depth measurement $d_\boldsymbol{i} = f_{\mathcal{R}_\mathcal{M}}(\boldsymbol{i})$ is obtained by scanning along the normal $\boldsymbol{n_i}$ for the zero level set of $\mathcal{T}$, i.e. $\mathcal{T}(\boldsymbol{m_i} + d_\boldsymbol{i} \cdot \boldsymbol{n_i}) = 0$. A unique characteristic of this ray casting technique is a surface model with high body coverage that allows to employ an efficient projective data association technique for closest point search. Given a point $\boldsymbol{p} \in \mathbb{R}^3$ we denote its projection onto $\mathcal{M}$ as $\widetilde{\boldsymbol{p}} = \mathcal{P}_\mathcal{M}(\boldsymbol{p})$, where $\mathcal{P}_\mathcal{M} : \mathbb{R}^3 \to \Omega_\mathcal{M}$ is the projection operator associated with the manifold $\mathcal{M}$. Consequently, we define the closest point of $\boldsymbol{p}$ on $\mathcal{S}_\mathcal{M}$ as:

$$\mathcal{CP}(\boldsymbol{p}, \mathcal{S}_\mathcal{M}) = \underset{\boldsymbol{x} \in \omega_{\widetilde{\boldsymbol{p}}}}{\operatorname{argmin}} \|\boldsymbol{p} - \boldsymbol{x}\|_2, \tag{7}$$

where $\omega_{\widetilde{\boldsymbol{p}}} \subset \Psi_\mathcal{M}$ denotes the local neighborhood of $f_{\mathcal{S}_\mathcal{M}}(\widetilde{\boldsymbol{p}})$ w.r.t. the surface topology defined by $\Omega_\mathcal{M}$. For the neighborhood search, we exploit this topology, thus superseding the need for any complex acceleration structures. Though our method is generic in the sense that arbitrary manifolds are supported, we found a manifold $\mathcal{M}$ based on a half cylinder as the most suitable representation in the context of real-time respiratory motion analysis. First, a half cylinder is a suitable approximation of the human thorax and, second, the projection of a point $\boldsymbol{p} \in \mathbb{R}^3$ onto the cylinder manifold $\mathcal{M}$ is given by a closed form solution.

2.3 Condition of RI based Joint Rigid and Non-rigid Registration

Without loss of generality, we formulate Eq. (1) in terms of concatenated vectors:

$$\mathcal{S}^k \equiv \overline{\boldsymbol{s}} = \overline{\boldsymbol{R}}^k(\overline{\boldsymbol{x}}^\mathcal{F} + \overline{\boldsymbol{u}}^k + (\overline{\boldsymbol{R}}^k)^{-1}\overline{\boldsymbol{t}}^k) = \overline{\boldsymbol{R}}^k\left(\overline{\boldsymbol{x}}^\mathcal{F} + \overline{\boldsymbol{u}}^k + \overline{\boldsymbol{v}}^k\right). \tag{8}$$

Here, $\overline{\boldsymbol{x}}^\mathcal{F} = \left[\boldsymbol{x}^\mathcal{F}_{0,x}, \boldsymbol{x}^\mathcal{F}_{0,y}, \boldsymbol{x}^\mathcal{F}_{0,z}, \ldots, \boldsymbol{x}^\mathcal{F}_{N,x}, \boldsymbol{x}^\mathcal{F}_{N,y}, \boldsymbol{x}^\mathcal{F}_{N,z}\right]^\top \in \mathbb{R}^{3N}$ concatenates the individual 3-D components (x, y, z) of the surface points $\boldsymbol{x}^\mathcal{F}_i$ in a single vector (similar $\overline{\boldsymbol{t}}^t$,$\overline{\boldsymbol{u}}^t$) and $\overline{\boldsymbol{R}}^t \in \mathbb{R}^{3N \times 3N}$ has the corresponding block structure. From Eq. (8)

it follows that the problem of RI based respiration analysis is ill-conditioned as, in general, $\overline{\boldsymbol{u}}^k$ and $\overline{\boldsymbol{v}}^k$ are not orthogonal or even linearly dependent. Hence, it is not always possible to separate rigid shifts from non-rigid respiratory motion. For example, for one single point $\boldsymbol{x}_i^{\mathcal{F}}$ and its corresponding point $\boldsymbol{x}_i^k \in \mathcal{S}^k$ at respiration state k, the displacement $\boldsymbol{d}_i^k = \boldsymbol{x}_i^k - \boldsymbol{x}_i^{\mathcal{F}}$ can not be traced back to rigid positioning ($\boldsymbol{d}_i^k \in \overline{\boldsymbol{v}}^k$) or non-rigid respiratory motion ($\boldsymbol{d}_i^k \in \overline{\boldsymbol{u}}^k$) as $\overline{\boldsymbol{v}}$ and $\overline{\boldsymbol{u}}^k$ are linearly dependent. In contrast, by using a higher body coverage, the individual displacement vectors $\overline{\boldsymbol{d}}^k \equiv \{\boldsymbol{d}_0^k, \ldots, \boldsymbol{d}_M^k\}$ with $M \leq N$ are spread across the whole body surface and feature different orientations and magnitudes. This facilitates linear independence of $\overline{\boldsymbol{v}}$ and $\overline{\boldsymbol{u}}^k$ and potentially allows for a better separation of $\overline{\boldsymbol{d}}^k$ into a rigid and a non-rigid component. We therefore introduce the condition metric $\mathcal{K}$ quantifying the orthogonality ($\mathcal{K} = 0$) or linear dependence ($\mathcal{K} = 1$) of a deformation field $\mathcal{U}^k \equiv \overline{\boldsymbol{u}}^k$ and rigid shifts as:

$$\mathcal{K}\left(\overline{\boldsymbol{u}}^k\right) = \frac{1}{3}\sum_{i=1}^{3} \frac{\left|\left\langle \overline{\boldsymbol{u}}^k, \overline{\boldsymbol{e}}_i \right\rangle\right|}{\left\|\overline{\boldsymbol{u}}^k\right\|_2 \left\|\overline{\boldsymbol{e}}_i\right\|_2} \in [0,1]. \tag{9}$$

Here, $\overline{\boldsymbol{e}}_1 = [1,0,0,\ldots,1,0,0] \in \mathbb{R}^{3N}$ denotes the vector corresponding to translations in x direction (and similar to $\overline{\boldsymbol{e}}_2$ and $\overline{\boldsymbol{e}}_3$). As we will show in the experiments, an increased body surface coverage using the proposed multi-sensor framework allows to decrease the condition metric $\mathcal{K}$ and results in a more reliable registration for patient alignment. Note that regions obtained by a higher body coverage may also include surface parts that do not move with respiration such as the arms and the head. However, such body parts can not be taken for granted if the reference surface was extracted from tomographic data (CT/MR) and we will focus on the more generic case of using a wide coverage thorax model.

2.4 Real-Time Respiration Analysis Using 4-D Shape Priors

The basic concept of respiratory motion analysis using 4-D shape priors is a respiration model $\mathcal{X}(\boldsymbol{b}) = \boldsymbol{\phi} + \boldsymbol{\Phi}\boldsymbol{b}$. The model mean $\boldsymbol{\phi}$ and the modes of variation $\boldsymbol{\Phi}$ are derived from a principal component analysis (PCA) of non-rigidly registered RI surfaces $\mathcal{S}^k$ acquired at different respiration states k [5]. The parameter vector $\boldsymbol{b}$ controls the model. The general work flow is an iterative model adaption based on point correspondences between an instantaneous surface $\mathcal{S}^{\mathcal{I}}$ and the i-th estimate for the deformable surface model $\mathcal{X}(\boldsymbol{b}^i)$ that is rigidly transformed by the i-th estimate of a rotation $\boldsymbol{R}^i$ and a translation $\boldsymbol{t}^i$. Respiration analysis using 4-D shape priors can benefit from a wide field field of view body coverage. This is because the iterative scheme explicitly relies on the joint rigid and non-rigid registration paradigm from Sect. 2.3 that is quantified by $\mathcal{K}$ in Eq. (9). Another issue concerns performance. The original work used an acceleration structure for efficient closest point search. However, as shown in later work [9], construction and query times are only acceptable for a small scale data size, in particular if a large number of iterations is required. In contrast, our framework does not require a complex acceleration structure but performs the closest point search by exploiting the projective data association technique described in Eq. (7).

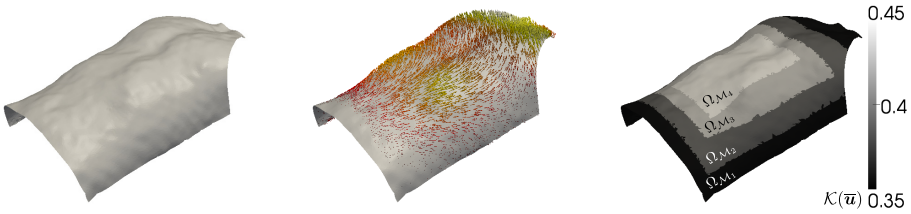

Fig. 2. Left to right: Reconstructed surface using manifold ray casting, estimated respiration deformation field $\mathcal{U}$ and its condition $\mathcal{K}(\overline{\boldsymbol{u}})$ for different surface regions $\Omega_{\mathcal{M}_i}$

3 Experiments and Results

Evaluation is performed with RI data from four male subjects using two Microsoft Kinect sensors (30 Hz, $\Omega_j = \mathbb{R}^{640\times480}$). For volumetric range data fusion we discretized $\mathcal{T}$ with 512^3 elements. Surface reconstruction using manifold ray casting to generate shapes $\mathcal{S}^k$ was then performed with $\Omega_{\mathcal{M}} = \mathbb{R}^{640\times480}$. Thus, $\left|\mathcal{S}^k\right| \sim 3 \cdot 10^5$. To account for sensor noise, we performed temporal averaging of the raw depth images $f_{\mathcal{R}_j}$, set the weighting of successive frames for temporal TSDF integration from Eq. (5) to $\alpha = 0.4$ and performed edge preserving filtering on $f_{\mathcal{R}_\mathcal{M}}$ as a final step. Fig. 2 depicts a surface that has been reconstructed using the proposed technique. Similar to [5], for each subject, we then extracted training shapes $\mathcal{S}^k$ covering one abdominal and one thoracic breathing cycle. The shapes were then non-rigidly registered using the *coherent point drift* algorithm [10]. PCA was applied to the registered shapes to set up the model $\mathcal{X}(\boldsymbol{b})$, with the number of modes chosen to cover 99% of the input variance. This resulted in 3 modes for all subjects. We then evaluated the condition $\mathcal{K}$ of a deformation field using Eq. (9) on four different surface regions $\Omega_{\mathcal{M}_4} \subset \Omega_{\mathcal{M}_3} \subset \Omega_{\mathcal{M}_2} \subset \Omega_{\mathcal{M}_1}$, see Fig. 2. This corresponds to different field of views where $\Omega_{\mathcal{M}_4}$ and $\Omega_{\mathcal{M}_3}$ simulate a single sensor setup whereas regions $\Omega_{\mathcal{M}_2}$ and $\Omega_{\mathcal{M}_1}$ require multiple sensors. For regions $\Omega_{\mathcal{M}_4}$ and $\Omega_{\mathcal{M}_1}$ the average condition metric $\mathcal{K}$ decreased from 0.34 to 0.27, 0.34 to 0.27, 0.38 to 0.29 and 0.39 to 0.32 for the four subjects. To investigate the sensitivity of respiratory motion analysis w.r.t. body coverage, we computed the respiration model $\mathcal{X}(\boldsymbol{b})$ for the four surface regions and performed a motion compensated patient alignment. Evaluation sequences covering one normal breathing cycle where captured in the same patient coordinate system as the training sequences. As the subjects did not move, the identity transformation was used to set up the ground truth. The average translation error Δt^k and rotation error Δr^k for evaluation frame k are then computed from the corresponding estimates t_i^k and r_i^k as $\Delta t^k = \frac{1}{3}\sum_{i=1}^{3} |t_i^k|$ and $\Delta r^k = \frac{1}{3}\sum_{i=1}^{3} |r_i^k|$, where the superscripts $i = 1, 2, 3$ denote the x,y and z axis. Fig. 3 depicts the individual transformation errors computed over all evaluation frames. We note that for subjects 1 and 4 the alignment error is rather high. This may be due to evaluation frames that are very different to the respiration states seen in the training phase. Yet, drifts are low compared to alignment without motion compensation that we evaluated for comparison. Regarding the influence of body coverage, we

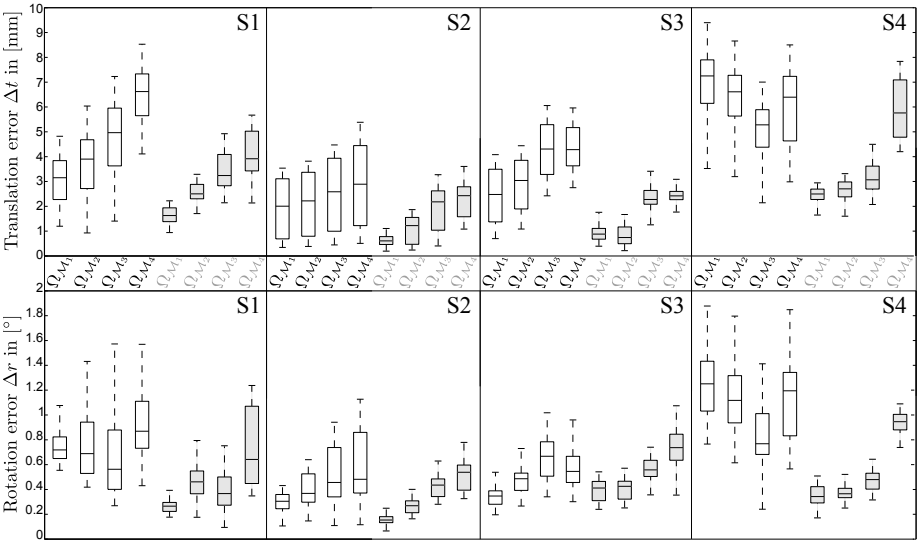

Fig. 3. Translation (top) and rotation (bottom) transformation estimation error for four male subjects S1-S4 and different surface regions $\Omega_{\mathcal{M}_4} \subset \Omega_{\mathcal{M}_3} \subset \Omega_{\mathcal{M}_2} \subset \Omega_{\mathcal{M}_1}$. Shaded boxes denote that motion compensation was performed.

state that in general the alignment error and its variance increases for smaller body surface regions. Over all subjects, a reduction of 3.65 mm to 1.38 mm and 0.73° to 0.30° for subregions $\Omega_{\mathcal{M}_4}$ and $\Omega_{\mathcal{M}_1}$ is noticeable. This corresponds to a factor of 2.7 and 2.4, respectively, underlining the benefit of a multi-camera system in RI based respiration analysis. We implemented our framework using CUDA on a PC running an Intel Core i7 3770K and an NVIDIA GTX 680. For the configuration described above, we compute a surface model $\mathcal{S}_{\mathcal{M}}$ in ~ 25 ms. Thus, our framework is beneficial for multi-camera systems with framerates of 20 fps [6]. For 4-D shape priors based deformation estimation with $|\mathcal{X}(\boldsymbol{b})| \sim 10^4$ model points and 100 iterations we achieve run times of 27 ms. However, in contrast to [5] our approach does not require a resampling of the surface $\mathcal{S}^{\mathcal{I}}$ and the closest point search uses all $3 \cdot 10^5$ surface points.

4 Conclusion

We presented a real-time framework for respiratory motion analysis using fused multi-view RI. As the major contribution, we proposed a volumetric approach to fuse RI data in conjunction with a ray casting technique to reconstruct a wide coverage surface model. Due to the high frame rates, our method is eligible to be used with clinical real-time multi-view RI systems. Our theoretical investigation and experiments further underline the benefit of a multi-sensor setup in model based respiratory motion analysis. This is promising w.r.t. subsequent steps such as internal/external motion inference that will be part of future research.

Acknowledgements This work was supported by the European Regional Development Fund and the Bayerisches Staatsministerium für Wirtschaft, Infrastruktur, Verkehr und Technologie under Grant No. IUK338/001.

References

1. McClelland, J., Hawkes, D., Schaeffter, T., King, A.: Respiratory motion models: A review. Med. Image Anal. 17(1), 19–42 (2013)
2. Fayad, H., Pan, T., Clement, J.F., Visvikis, D.: Technical note: Correlation of respiratory motion between external patient surface and internal anatomical landmarks. Med. Phys. 38(6), 3157–3164 (2011)
3. Bauer, S., Berkels, B., Ettl, S., Arold, O., Hornegger, J., Rumpf, M.: Marker-Less Reconstruction of Dense 4-D Surface Motion Fields Using Active Laser Triangulation for Respiratory Motion Management. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) MICCAI 2012, Part I. LNCS, vol. 7510, pp. 414–421. Springer, Heidelberg (2012)
4. Schaerer, J., Fassi, A., Riboldi, M., Cerveri, P., Baroni, G., Sarrut, D.: Multidimensional respiratory motion tracking from markerless opticalsurface imaging based on deformable mesh registration. Phys. Med. Biol. 57(2), 357–373 (2012)
5. Wasza, J., Bauer, S., Hornegger, J.: Real-time Motion Compensated Patient Positioning and Non-rigid Deformation Estimation Using 4-D Shape Priors. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) MICCAI 2012, Part II. LNCS, vol. 7511, pp. 576–583. Springer, Heidelberg (2012)
6. Price, G.J., Parkhurst, J.M., Sharrock, P.J., Moore, C.J.: Real-time optical measurement of the dynamic body surface for use in guided radiotherapy. Phys. Med. Biol. 57(2), 415 (2012)
7. Curless, B., Levoy, M.: A volumetric method for building complex models from range images. In: Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH 1996, pp. 303–312. ACM (1996)
8. Newcombe, R.A., Davison, A.J., Izadi, S., Kohli, P., Hilliges, O., Shotton, J., Molyneaux, D., Hodges, S., Kim, D., Fitzgibbon, A.: KinectFusion: Real-time dense surface mapping and tracking. In: 2011 10th IEEE International Symposium on Mixed and Augmented Reality, pp. 127–136. IEEE (2011)
9. Bauer, S., Wasza, J., Lugauer, F., Neumann, D., Hornegger, J.: Real-Time RGB-D Mapping and 3-D Modeling on the GPU Using the Random Ball Cover. In: Fossati, A., Gall, J., Grabner, H., Ren, X., Konolige, K. (eds.) Consumer Depth Cameras for Computer Vision - Research Topics and Applications, London, UK. Advances in Computer Vision and Pattern Recognition, pp. 27–48 (2013)
10. Myronenko, A., Song, X.: Point set registration: Coherent point drift. IEEE Trans. Pattern Anal. Mach. Intell. 32(12), 2262–2275 (2010)