Real-Time Respiratory Motion Analysis Using 4-D Shape Priors

Jakob Wasza, Peter Fischer, *Student Member, IEEE*, Heike Leutheuser *Student Member, IEEE EMBS*, Tobias Oefner, Christoph Bert, Andreas Maier*, *Member, IEEE*, Joachim Hornegger, *Member, IEEE*

***Abstract*—Respiratory motion analysis based on range imaging (RI) has emerged as a popular means of generating respiration surrogates to guide motion management strategies in computer assisted interventions. However, existing approaches employ heuristics, require substantial manual interaction, or yield highly redundant information. In this paper we propose a framework that uses *pre-procedurally* obtained 4-D shape priors from patient-specific breathing patterns to drive *intra-procedural* RI-based real-time respiratory motion analysis. As the first contribution, we present a shape motion model enabling an unsupervised decomposition of respiration induced high-dimensional body surface displacement fields into a low-dimensional representation encoding thoracic and abdominal breathing. Second, we propose a method designed for GPU architectures to quickly and robustly align our models to high coverage multi-view RI body surface data. With our fully automatic method we obtain respiration surrogates yielding a Pearson correlation coefficient (PCC) of 0.98 with conventional surrogates based on manually selected regions on RI body surface data. Compared to impedance pneumography as a respiration signal that measures the change of lung volume, we obtain a PCC of 0.96. Using off-the-shelf hardware, our framework enables high temporal resolution respiration analysis at 50 Hz.**

***Index Terms*—Radiation Therapy, Respiratory Motion, Range Imaging, Graphics Processing Unit (GPU)**

I. Introduction

Respiratory motion is a major issue in computer assisted interventions such as image guided radiation therapy [1], [2]. One option for motion mitigation is to reduce the respiration amplitude using active breath-control systems or abdominal pressure [3]. However, such techniques are often not practicable and advanced methods such as beam tracking or beam gating are based on continuous respiratory motion monitoring and analysis [4], [5], [6]. For this purpose, external respiration surrogates computed from sparsely distributed reflective infrared (IR) markers attached to the body or dense range imaging (RI) to capture the entire body surface have gained increasing popularity. Especially in combination with patient-specific motion models that correlate external surrogates to internal tumor movement [7], [8], the advantages of IR markers or RI sensors, namely their real-time capability and absence of ionizing radiation, can be exploited. Clinically available IR tracking solutions include the Synchrony system (Accuray Inc., Sunnyvale, USA) and the ExacTrac module (Brainlab AG, Feldkirchen, Germany). For RI solutions, the AlignRT system (Vision RT, London, UK) is clinically available and novel RI sensor concepts to tackle the limitations of existing principles have been recently presented [9], [10], [11].

One advantage of RI solutions compared to IR tracking methods is that they do not require auxiliary equipment such as reflective vests or markers attached to the patient. However, with RI sensors, a respiration surrogate is not given per se and various approaches have been proposed.

A. Related work

Commonly, RI based external respiration surrogate signals are derived from the movement of body surface regions [12], [13], [14], [15], [16]. One issue with such methods is that the surrogates are based on variations of 3-D points or plain 1-D depth measurements rather than true 3-D surface deformations. Further, the number of body surface regions, their size and position must be manually and repeatedly chosen. This is a heuristic approach that is potentially error prone due to low reproducibility of landmark selection and requires substantial manual interaction which complicates clinical workflows.

Alternative approaches to overcome these limitations compute external body deformations using non-rigid surface registration techniques [9], [17], [18]. By design, these methods allow anatomically plausible 3-D surface landmark tracking and are capable of avoiding recurring manual landmark selection by registering pre-procedurally labeled planning data to intra-procedurally acquired RI body surface data. The major advantage of non-rigid registration methods, however, is the ability to generate high-dimensional respiration surrogates in a fully automatic manner via dense surface displacement fields. An issue with such high-dimensional surrogates is that they contain redundant information and the low-dimensional non-redundant factors must be identified. Approaches to address this problem are to calculate the mean displacement magnitude or the movement along the principal surface deformation

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. The authors gratefully acknowledge funding by the Erlangen Graduate School in Advanced Optical Technologies (SAOT) by the German Research Foundation (DFG) in the framework of the German excellence initiative.

J. Wasza, P. Fischer, H. Leutheuser, T. Oefner, A. Maier* and J. Hornegger are with the Pattern Recognition Lab, Department of Computer Science, Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), Martensstr. 3, 91058 Erlangen, Germany (e-mail: andreas.maier@fau.de).

C. Bert is with the Department of Radiation Oncology, Universitätsklinikum Erlangen and the Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), Germany.

P. Fischer, A. Maier and J. Hornegger are with the Erlangen Graduate School in Advanced Optical Technologies (SAOT).

direction [17]. However, such reduction schemes do not account for the anatomical source of motion. Besides these methodological challenges, as an important practical aspect, real-time capability is an open issue with non-rigid surface registration.

B. Contributions

In this paper we propose a framework for RI-based respiratory motion analysis to address the limitations of existing approaches. Our framework is based on prior knowledge of patient-specific 4-D shape deformations. This information is encoded in a surface motion model that is *pre-procedurally* trained from non-rigidly registered 3-D body surface data acquired at different respiration states. *Intra-procedurally*, the motion model is registered to the patient's current body surface acquired with multi-view RI. The intrinsic parameters that govern the registered motion model then define a non-redundant external respiration surrogate. Further, our framework supports the reconstruction of respiration induced dense body surface displacement fields to generate custom respiration surrogates. A sketch of our proposed framework is given in Fig. 1.

The feasibility of using prior shape knowledge for respiration management was demonstrated in our preceding work for motion compensated patient positioning [19], [20]. In this paper, we substantially extend this idea for real-time respiratory motion analysis by the following two main contributions: first, we propose a sparse motion model that enables an unsupervised decomposition of respiration induced high-dimensional dense surface displacement fields into a thoracic and abdominal component. Second, we present a fast and robust registration technique to align motion models to high coverage body surface data computed from multi-view RI. To cope with real-time constraints, our framework is designed for general purpose graphics processing unit (GPU) architectures.

II. Methods

In the *pre-procedural* training phase we denote $\mathcal{P} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\},\ \boldsymbol{x}_n \in \mathbb{R}^3$ as the point set that describes a body surface. Without loss of generality, the point set $\mathcal{P}$ can be equivalently described as a vector $\boldsymbol{p} \in \mathbb{R}^{3N}$ according to $\mathcal{P} \equiv \boldsymbol{p} = ((\boldsymbol{x}_1)^\top, \ldots, (\boldsymbol{x}_N)^\top)^\top$. Using point sets instead of closed surfaces enables a generic framework that is independent from the acquisition modality to capture prior knowledge on 4-D shape deformations. For example, tissue-to-air thresholding techniques to extract body surfaces from tomographic planning data such as computed tomography (CT) or magnetic resonance imaging (MRI) may exhibit a different topology and sampling density compared to RI-based surfaces. Further, laser scanners or markers do not provide a continuous surface at all.

In the *intra-procedural* respiratory motion analysis phase we denote an RI in a multi-view setup associated with the k-th sensor as $\mathcal{R}_k : \Omega_k \to \mathbb{R}^+$. $\mathcal{R}_k$ is sampled on the domain $\Omega_k \subset \mathbb{R}^2$ with $N_1 \times N_2$ 1-D orthogonal range measurements $r_{k,\boldsymbol{i}}$ in [mm] with $\boldsymbol{i} \in \Omega$ that correspond to the depth of a 3-D point $\boldsymbol{x} \in \mathbb{R}^3$ along the viewing axis of the RI camera. Consistently, we denote the 3-D surface corresponding to $\mathcal{R}_k$ as $\mathcal{S}_k : \Omega_k \to \mathbb{R}^3$. We further denote $P_k : \mathbb{R}^3 \to \Omega_k$ as the operator that projects the point $\boldsymbol{x}$ into Ω_k. For a surface point $\boldsymbol{x}$ captured by the k-th sensor we thus have:

$$\begin{aligned} \mathcal{S}_k(P_k(\boldsymbol{x})) &= \boldsymbol{R}_k^\top \left(\boldsymbol{x} - \boldsymbol{t}_k\right), \\ \mathcal{R}_k(P_k(\boldsymbol{x})) &= r_{k,\boldsymbol{i}} = (0,0,1) \cdot \mathcal{S}_k(P_k(\boldsymbol{x})), \end{aligned} \tag{1}$$

where $\boldsymbol{R}_k \in \mathrm{SO}(3)$ and $\boldsymbol{t}_k \in \mathbb{R}^3$ denote the rotation and translation to define the k-th RI sensor coordinate system. Further, surface normals $\mathcal{N}_k : \Omega_k \to \mathbb{R}^3$ can be readily computed from $\mathcal{S}_k$ using for example difference quotients.

For a sketch of the different phases, steps and data of our framework see Fig. 1.

A. Pre-procedural 4-D Shape Motion Models

We denote $\mathcal{P}^0 = \{\boldsymbol{x}_1^0, \ldots, \boldsymbol{x}_N^0\}$ as the point set that describes the body surface at a *fixed* reference respiration state e. g. fully exhale. Analogously, $\mathcal{P}^s = \{\boldsymbol{x}_1^s, \ldots, \boldsymbol{x}_N^s\}$ is the point set denoting the *moving* body surface at respiration state $s = 1, \ldots, S$. The moving surface points are obtained by warping the fixed surface points with an elastic deformation. This deformation is represented by the displacement field $\mathcal{U}^s = \{\boldsymbol{u}_1^s, \ldots, \boldsymbol{u}_N^s\},\ \boldsymbol{u}_n^s \in \mathbb{R}^3$ encoding the point-wise displacements induced by respiratory motion as:

$$\boldsymbol{x}_n^s = \boldsymbol{x}_n^0 + \boldsymbol{u}_n^s,\ \forall \boldsymbol{x}_n^s \in \mathcal{P}^s. \tag{2}$$

Depending on the acquisition modality, the displacement fields $\mathcal{U}^s$ can be computed using dedicated RI surface matching techniques [9], [17], volumetric registration for CT/MRI and surfaces that are represented implicitly by distance transforms, or point-set alignment methods [21].

1) Motion Model Training: Training a 4-D shape motion model is to recover the intrinsic structure that governs the high-dimensional representations of the training data $\mathcal{T} = \{\boldsymbol{p}^1, \ldots, \boldsymbol{p}^S\}$ with $\boldsymbol{p}^s \equiv \mathcal{P}^s$. For this purpose, we use *Principal Component Analysis* (PCA) as a linear dimensionality reduction technique. We favor PCA over non-linear methods such as *Manifold Learning* as the latter typically requires a large number of training samples which is prohibitive if the training shapes are extracted from CT or MRI data. Further, non-linear dimensionality reduction techniques do not offer advantages per-se as we showed in preceding work [22].

One interpretation of PCA is to find an orthonormal basis that allows to reconstruct the training data $\mathcal{T}$ in a least squares sense [23]. This is readily done by an eigendecomposition of the covariance matrix $\boldsymbol{C}$ of the mean-centered training data:

$$\boldsymbol{C} = \frac{1}{S}\sum_{s=1}^{S}(\boldsymbol{p}^s - \overline{\boldsymbol{p}})(\boldsymbol{p}^s - \overline{\boldsymbol{p}})^\top,\ \overline{\boldsymbol{p}} = \frac{1}{S}\sum_{s=1}^{S}\boldsymbol{p}^s. \tag{3}$$

Let $\boldsymbol{P} \in \mathbb{R}^{3N \times L}$ denote the PCA basis that column-wise contains the *modes of variation* defined by the eigenvectors $\boldsymbol{e}_l \in \mathbb{R}^{3N}$ corresponding to the L largest eigenvalues λ_l of $\boldsymbol{C}$. In practice, L is chosen to match the intrinsic dimensionality of the training data. Using a cumulative variance criterion this translates to $\sum_{l=1}^{L}\lambda_l \geq \delta \sum_{s=1}^{S}\lambda_s$, i. e. the model accounts for more than $(100 \cdot \delta)\%$ of the variance in the input training

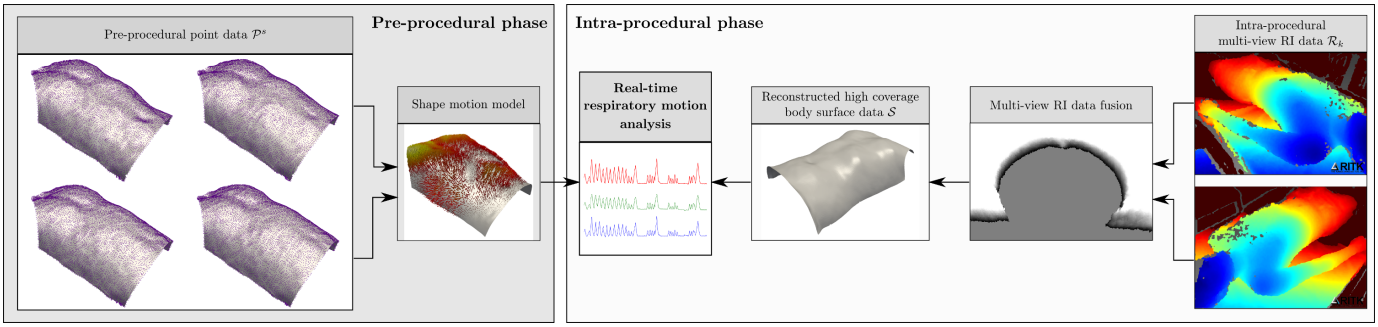
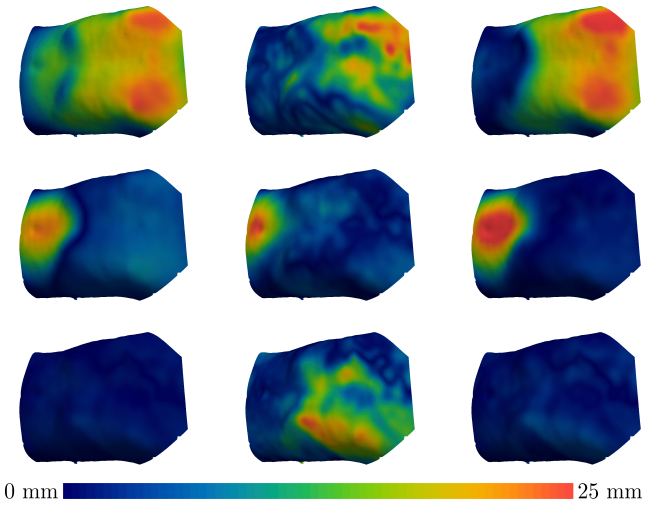


Fig. 1. The respiratory motion analysis framework as proposed in this work. The method is divided into a pre-procedural and an intra-procedural phase. In the pre-procedural phase, a shape motion model is generated from prior knowledge of 4-D shape deformations. In the intra-procedural phase, the motion model is registered to RI surface data yielding a respiration surrogate.



(a) Standard PCA. (b) Conventional VR. (c) Weighted VR.

Fig. 2. Exemplary depiction of the three leading modes of variation $\boldsymbol{e}_l$ (top to bottom) for different motion models. The magnitude of the individual surface displacements in [mm] is color coded. The proposed weighted VR model allows for a differentiation between thoracic and abdominal breathing.

data. For a test point set $\mathcal{P} \equiv \boldsymbol{p} \in \mathbb{R}^{3N}$ its approximation w. r. t. the PCA basis $\boldsymbol{P}$ is traditionally computed as:

$$\boldsymbol{p} \approx \boldsymbol{p}^{\boldsymbol{b}} = \overline{\boldsymbol{p}} + \boldsymbol{P}\left(\boldsymbol{P}^{\top}(\boldsymbol{p} - \overline{\boldsymbol{p}})\right) = \overline{\boldsymbol{p}} + \boldsymbol{P}\boldsymbol{b}, \tag{4}$$

where the model parameter vector $\boldsymbol{b} = [b_1, \ldots, b_L] \in \mathbb{R}^L$ is the low-dimensional representation of $\boldsymbol{p}$.

Essentially, $\boldsymbol{b}$ corresponds to a least squares estimate with regard to point-wise Euclidean distances between corresponding points in the model and the test point set. As our framework is based on the registration of the motion model to instantaneous RI surface data, this estimator is potentially prone to noise and outliers that are typical with range imaging. A robust method to account for this issue is investigated in Section II-C.

Valid shapes are restricted to $b_l \in [-3\sqrt{\lambda_l}, +3\sqrt{\lambda_l}]$ in order to enforce plausibility w. r. t. to the training data. This interval assumes that the training data is normally distributed which is not necessarily true for respiration related samples. However, in practice, we found these values as reasonable bounds.

2) Model-based Respiration Surrogates: For respiratory motion analysis, we exploit the fact that the modes of variation $\boldsymbol{e}_l$ contained in $\boldsymbol{P}$ encode non-rigid body surface displacement fields, see Fig. 2a for an illustration. Thus, the individual model parameters b_l constitute respiration surrogates that describe the change of surface extent w. r. t. the l-th mode relative to the mean shape $\overline{\boldsymbol{p}}$.

In this work we assume that maximal values of the parameters $\boldsymbol{b}_l$ correspond to the state of full inhale. This property is not given per se as the modes of variation $\boldsymbol{e}_l$ computed by PCA have arbitrary signs. One automatic approach to this problem is to flip the signs such that the model instance $\boldsymbol{p}^{\boldsymbol{b}}$ with maximum extent is obtained.

Following the model plausibility criterion from the preceding section, we shift the model parameters according to $b'_l = b_l + 3\sqrt{\lambda_l}$. This enforces the parameters b'_l to be positive and to describe the surface extent relative to the most compact or fully exhale shape allowed by the model. Consequently, we compute the respiration surrogate $\sigma_l \in \mathbb{R}^+$ for the l-th mode and the joint surrogate $\sigma_{\mathrm{J}} \in \mathbb{R}^+$ across all modes as:

$$\begin{aligned} \sigma_l &= b'_l, \\ \sigma_{\mathrm{J}} &= \|\boldsymbol{b}'\|_2. \end{aligned} \tag{5}$$

Note that (5) is not valid for raw parameters $b_l \in \mathbb{R}$ as a potential negative sign would be lost.

3) Sparse Motion Models: One inherent problem with PCA is that the modes $\boldsymbol{e}_l$ are defined to yield a minimum reconstruction error w. r. t. the training samples $\boldsymbol{p}^s$ [23]. Thus, the individual components of $\boldsymbol{e}_l$ that are also known as *loadings* in general do not exhibit sparsity. As a consequence, the modes $\boldsymbol{e}_l$ represent global abstract deformations that do not allow for a respiration analysis w. r. t. to anatomically plausible distinct breathing patterns, see Fig. 2a for an illustration.

In a preliminary study we have shown that so-called *Varimax Rotations* (VR) known from factor analysis [24] allow to compute sparse respiratory motion modes [25]. However, this study was based on plain 1-D depth data which is a simplifying assumption that does not reflect complex 3-D surface deformations. In fact, we found that a naive application of VRs for complex 3-D deformation data further hinders an intuitive interpretation. This issue is depicted in Fig. 2b where the individual modes exhibit sparse but widespread

deformation components. The reason for this can be traced back to the theoretical concepts of factor rotations.

The general principle behind VR is to find a rotation $\boldsymbol{R} \in \mathrm{SO}(L)$ such that the modes of variation $\boldsymbol{e}_l^{\mathrm{VR}}$ contained column-wise in the rotated basis $\boldsymbol{P}^{\mathrm{VR}} = \boldsymbol{P}\boldsymbol{R}$ exhibit a more simple structure. The sought rotation $\boldsymbol{R}$ must maximize [24]:

$$\mathcal{J}_{\boldsymbol{R}}^{\mathrm{VR}} = \sum_{l=1}^{L} \left(\sum_{q=1}^{3N} [\boldsymbol{P}\boldsymbol{R}]_{q,l}^4 - \left(\sum_{q=1}^{3N} [\boldsymbol{P}\boldsymbol{R}]_{q,l}^2 \right)^2 \right). \quad (6)$$

Using $\mathrm{Var}(X) = \mathrm{E}(X^2) - [\mathrm{E}(X)]^2$ with $\mathrm{Var}(X)$ denoting the variance and $\mathrm{E}(X)$ the expected value of a random variable X, (6) corresponds to maximizing the variances of *squared* loadings, i. e. $X \propto [\boldsymbol{P}\boldsymbol{R}]_{q,l}^2$, across the modes of variation $\boldsymbol{e}_l^{\mathrm{VR}}$ contained in $\boldsymbol{P}^{\mathrm{VR}}$. Due to orthonormality constraints on $\boldsymbol{R}$, the only possibility to increase the mode-wise variances in (6) is to bring some loadings close to zero and make others grow large [26].

The problem with this formulation is that it rigorously favors sparsity w. r. t. the *scalar*-valued components in the *normalized* modes $\boldsymbol{e}_l^{\mathrm{VR}}$. Thus, the *vector*-valued nature of point-wise 3-D displacements as well as the variances λ_l as a measure of respiration magnitude is lost.

We address this issue by using a weighted VR (WVR) model extending the standard VR optimization problem from (6) with a diagonal weighting matrix Λ according to:

$$\mathcal{J}_{\boldsymbol{R}}^{\mathrm{WVR}} = \sum_{l=1}^{L} \left(\sum_{q=1}^{3N} [\boldsymbol{P}\Lambda\boldsymbol{R}]_{q,l}^4 - \left(\sum_{q=1}^{3N} [\boldsymbol{P}\Lambda\boldsymbol{R}]_{q,l}^2 \right)^2 \right). \quad (7)$$

For $\Lambda = \mathrm{diag}(1, \ldots, 1)$ the standard VR principle is retained whereas for $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_L)$ the solution is similar to conventional PCA with the difference that the variance of the *squared* loadings weighted by λ_l is maximized. In this work, we use the mode-specific standard deviations defined by the eigenvalues λ_l as weighting factors, i. e. $\Lambda = \mathrm{diag}(\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_L})$. The rationale behind this scheme is that these eigenvalues are directly related to the amount of shape variation in the training set that by definition reflects different respiration states with varying surface deformation. Thus, sparsity w. r. t. the respiration magnitude is injected in our WVR model.

The optimization problems from (6) and (7) can be solved by using an iterative scheme based on singular value decompositions [26]. One issue with the WVR model is that the basis $\boldsymbol{P}^{\mathrm{WVR}} = \boldsymbol{P}\Lambda\boldsymbol{R}^{\mathrm{WVR}}$ is in general not orthogonal. Here, $\boldsymbol{R}^{\mathrm{WVR}}$ denotes the solution from (7). We address this issue by estimating the rotation $\boldsymbol{R}^{\mathrm{WVR}}$ with the WVR approach from (7) but compute the WVR basis *without* the weighting matrix Λ analogous to the standard VR approach according to $\boldsymbol{P}^{\mathrm{WVR}} = \boldsymbol{P}\boldsymbol{R}^{\mathrm{WVR}}$. The modes that are obtained with our WVR scheme are exemplarily shown in Fig. 2c.

B. Intra-procedural Multi-View RI Body Surface Models

Our respiratory motion framework is based on the registration of a potentially high coverage *pre-procedurally* trained motion model to instantaneous RI body surface data. Thus, the RI body surface data ideally should exhibit high coverage, too. Due to the limited field-of-view of a single camera multiple RI sensors are advantageous to obtain this property.

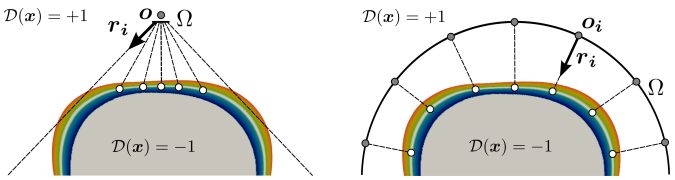


Fig. 3. Surface reconstruction using conventional pinhole (left) and manifold (right) ray casting of a distance transform $\mathcal{D}$. Color-coded is the support region ϵ, i. e. $\mathcal{D}(\boldsymbol{x}) \in (-1, +1)$. Note that the diverging rays with a pinhole camera model do not allow for high coverage surface models.

A real-time capable method for fusing the multi-view RI data and reconstructing a unifying surface representation was proposed in our prior work [20]. Due to the importance of this scheme for the proposed robust and fast motion model registration in Section II-C we now outline the basic working principles.

1) Data Fusion Using Distance Transforms: We represent the surface data $\mathcal{S}_k$ implicitly via its corresponding distance transform $\mathcal{D}_k : \Gamma_k \rightarrow [-1, +1]$ that holds the distance to the closest point on the RI surface $\mathcal{S}_k$. $\Gamma \subset \mathbb{R}^3$ denotes a suitable region embedding $\mathcal{S}_k$. Similar to Curless and Levoy [27] the distance transform is approximated based on the relationship from (1) in a support region around $\mathcal{S}_k$ as:

$$\mathcal{D}_k(\boldsymbol{x}) = \eta \left(\mathcal{R}_k \left(P_k \left(\boldsymbol{x} \right) \right) - (0, 0, 1) \cdot \boldsymbol{R}_k^{\top} \left(\boldsymbol{x} - \boldsymbol{t}_k \right) \right), \quad (8)$$

where $\eta(d) = \min(1, |\epsilon^{-1} d|) \cdot \mathrm{sign}(d)$ accounts for the support region controlled by $\epsilon \in \mathbb{R}^+$. The fused implicit representation $\mathcal{D}$ is then given by a weighted average of the individual distance transforms $\mathcal{D}_k$ [20].

2) Surface Reconstruction Using Ray Casting: For reconstructing an explicit surface $\mathcal{S}$ from the fused implicit representation $\mathcal{D}$ we use a *Manifold Ray Casting* technique introduced in our previous work [20]. This approach overcomes the limited body surface coverage with conventional ray casting, see Fig. 3 for an illustration.

Mathematically, we simulate an RI sensor by estimating the 1-D depth measurements $\mathcal{R}$ such that the corresponding 3-D surface points $\mathcal{S}$ coincide with the zero level-set of the distance transform $\mathcal{D}$, i. e.:

$$\mathcal{D}\left(\boldsymbol{R} \left(\boldsymbol{o_i} + \mathcal{R}(\boldsymbol{i}) \boldsymbol{r_i} \right) + \boldsymbol{t} \right) \stackrel{!}{=} 0, \ \forall \boldsymbol{i} \in \Omega, \quad (9)$$

where $\boldsymbol{R} \in \mathrm{SO}(3)$ and $\boldsymbol{t} \in \mathbb{R}^3$ similar to (1) account for the virtual sensor position and orientation. Further, $\boldsymbol{o_i} \in \mathbb{R}^3$ and $\boldsymbol{r_i} \in \mathbb{R}^3$ with $\|\boldsymbol{r_i}\|_2 = 1$ denote the starting position and direction of the ray associated with $\boldsymbol{i}$, respectively. Both $\boldsymbol{o_i}$ and $\boldsymbol{r_i}$ are computed from the manifold and basically constitute the counterpart of the projection operator from (1). Thus, there also exists a projection operator $P : \mathbb{R}^3 \rightarrow \Omega$.

As a prerequisite, the virtual RI domain Ω must exhibit a 2-D parameterization. In this work we use a manifold based on a half-cylinder as a coarse approximation of the human torso. Thus, a 2-D parameterization of Ω is readily obtained via cylindrical coordinates.

C. Intra-procedural Motion Model Registration

In this paper, we assume that the patient is aligned w. r. t. the *pre-procedural* planning data. For existing approaches we refer to our prior work on motion-compensated alignment [19], [20] or the motion-gated patient setup by Placht et al. [28].

For an aligned patient, we have to estimate the model parameter $\boldsymbol{b}$ such that the corresponding model instance $\mathcal{P}^{\boldsymbol{b}}$ fits the instantaneous RI surface data $\mathcal{S}$. However, directly using the least squares estimator from (4) is prohibitive. First, correspondences between the individual motion model components and RI surface points are not known. Second, the least squares estimator that drives the computation of $\boldsymbol{b}$ is not robust to outliers caused by erroneous correspondences, partial surface matching issues and noise in the RI data.

1) Robust Alignment: A technique that inherently addresses these issues is the *Coherent Point Drift* (CPD) method [21]. The CPD framework models a point set registration as a probability density estimation problem with one point set representing the centroids of a *Gaussian Mixture Model* (GMM) and the second point set representing the data points. The registration is then performed by re-parameterizing the GMM centroids to minimize the negative Log-Likelihood of the data points. This GMM based alignment scheme is solved by the Expectation Maximization (EM) algorithm, eventually leading to an iterative weighted multi-link registration problem [21].

In contrast to the original CPD formulation, we do not quantify the alignment of the model instance $\mathcal{P}^{\boldsymbol{b}}$ and the instantaneous RI surface $\mathcal{S}$ based on *point-to-point* distances but instead use a *point-to-tangent* metric. This is motivated by the fact that the latter has shown superior performance in terms of attraction range and convergence [29]. Our CPD-like iterative optimization problem to be minimized is:

$$\begin{aligned}\mathcal{J}^{\text{CPD}}_{\boldsymbol{b},\sigma} = &\sum_{n=1}^{N}\sum_{m=1}^{|\mathcal{S}|} p(\boldsymbol{x}_n^{\boldsymbol{b}}|\boldsymbol{y}_m)\left(\frac{(\boldsymbol{x}_n^{\boldsymbol{b}}-\boldsymbol{y}_m)\cdot\boldsymbol{n}_m}{\sigma}\right)^2 \\ &+\sum_{n=1}^{N}\sum_{m=1}^{|\mathcal{S}|} p(\boldsymbol{x}_n^{\boldsymbol{b}}|\boldsymbol{y}_m)\log(\sigma^2),\end{aligned} \tag{10}$$

with GMM centroids $\boldsymbol{x}_n^{\boldsymbol{b}} \in \mathcal{P}^{\boldsymbol{b}} \equiv \boldsymbol{p}^{\boldsymbol{b}} = \overline{\boldsymbol{p}} + \boldsymbol{P}\boldsymbol{b}$ similar to (4). Further, $\boldsymbol{y}_m \in \mathcal{S}$ denote the data points with associated surface normals $\boldsymbol{n}_m \in \mathcal{N}$. The posterior probabilities $p(\boldsymbol{x}_n^{\boldsymbol{b}}|\boldsymbol{y}_m)$ of GMM centroids given the data points are computed using the previous parameter estimates according to:

$$p(\boldsymbol{x}_n^{\boldsymbol{b}}|\boldsymbol{y}_m) = \frac{\exp\left(-\left(\frac{\left(\boldsymbol{x}_n^{\boldsymbol{b}'}-\boldsymbol{y}_m\right)\cdot\boldsymbol{n}_m}{\sigma'}\right)^2\right)}{\sum_{k=1}^{N}\exp\left(-\left(\frac{\left(\boldsymbol{x}_k^{\boldsymbol{b}'}-\boldsymbol{y}_m\right)\cdot\boldsymbol{n}_m}{\sigma'}\right)^2\right)+c}, \tag{11}$$

where $c = \sqrt{2\pi\sigma'^2}\frac{w}{1-w}\frac{N}{|\mathcal{S}|}$ with $w \in [0,1]$ to account for noise and outliers. Symbols marked with a prime denote the previous estimate [21]. As shown in Appendix A, minimizing (10) w. r. t. $\boldsymbol{b}$ corresponds to a linear equation system that can be solved efficiently by using *Cholesky* decompositions.

2) Fast CPD Approximation: A crucial issue with the CPD formulation is run-time performance. Though the original work reported significant speed-ups using the fast Gauss transform [21], we found the CPD registration scheme computationally too expensive for the real-time scenario at hand.

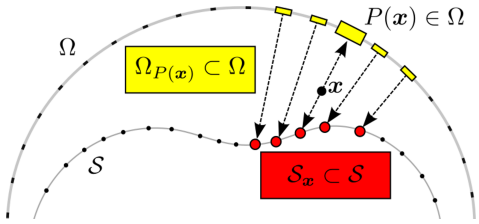


Fig. 4. Surface neighborhood computation using projective data association. For a test point $\boldsymbol{x}$ its closest points $\mathcal{S}_{\boldsymbol{x}} \subset \mathcal{S}$ are given via the local neighborhood $\Omega_{P(\boldsymbol{x})} \subset \Omega$ around its projection $P(\boldsymbol{x}) \in \Omega$ into the RI sensor domain Ω, cf. (1).

Inspired by the original work that also used truncated Gaussians [21] we thus propose an approximative scheme tailored to RI data: for a given model point $\boldsymbol{x}_n^{\boldsymbol{b}}$ we do not use all points $\boldsymbol{y}_m \in \mathcal{S}$ in (10) but only a small subset of closest points $\boldsymbol{y}_{n,m} \in \mathcal{S}_{\boldsymbol{x}_n^{\boldsymbol{b}}} \subset \mathcal{S}$, see Fig. 4. Similar to [21] we argue that this is a valid approximation as the weights defined by the posteriors and the kernel width rapidly decay with increasing distance. The key step towards real-time computation is that for a point $\boldsymbol{x}_n^{\boldsymbol{b}}$ its closest neighbors $\boldsymbol{y}_{n,m} \in \mathcal{S}_{\boldsymbol{x}_n^{\boldsymbol{b}}}$ can be computed efficiently using a projective data association (PDA) scheme that is inherent to our manifold ray casting technique outlined in Section II-B2.

The working principle of this multi-link point correspondences approach is illustrated in Fig. 4 and the mathematical formulation for the final approximated CPD alignment problem as well as implementation aspects for real-time computation are given in Appendix A.

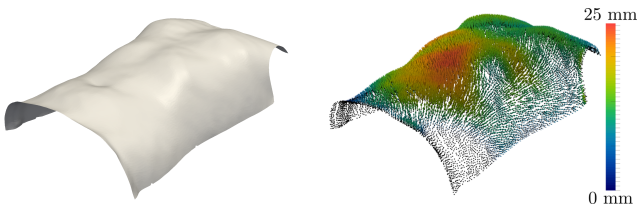
III. Experiments and Results

The experiments in this paper are concerned with two aspects. First, we perform a thorough assessment of the methodology inherent to our respiratory motion analysis framework. Second, we compare the respiratory motion surrogates obtained with our model-based approach to a respiration sensor that is independent of surface motion.

A. Assessment of Methodology

The purpose of this experiment is to assess the general aspects of the proposed respiratory motion analysis framework regarding (i) the ability of motion models to adapt to unseen surface data, (ii) the robustness of model registration in the presences of outliers in RI data, (iii) the suitability of motion models to generate low-dimensional respiration surrogates and (iv) the run-time performance.

1) Setup and Data: We captured RI data from four male subjects $S_1, \ldots, S_4$ using two Microsoft Kinect RI sensors ($\sim$ 30 Hz, $\Omega_{1,2} = \mathbb{R}^{640\times480}$) with an acquisition distance of $\sim$ 1 m. For RI data fusion we discretized the distance transform as $\Gamma = \mathbb{R}^{256\times256\times256}$ and the manifold domain for surface reconstruction as $\Omega = \mathbb{R}^{640\times480}$. To account for missing data and noise we performed post-processing using normalized convolution and guided filtering similar to [30].



(a) RI surface $\mathcal{S}^s$. (b) Point set $\mathcal{P}^0$ with $\mathcal{U}^s$.

Fig. 5. Exemplary depiction of a cropped and post-processed reference RI surface $\mathcal{S}^s$ (left) and the training point set $\mathcal{P}^0$ denoted as black dots with the corresponding displacement field $\mathcal{U}^s$ (right) color coded by the magnitude.

For training our 4-D motion model, the subjects were instructed to perform one cycle of thoracic and abdominal breathing, respectively. Each cycle was sampled with 1 Hz which corresponds to $5-7$ different respiration states for the individual sequences, see Table IV in Appendix B for a detailed listing. We purposely employed a sparse temporal sampling to stress the fact that our method is generic in a sense that no specific modality to capture surface data from different respiration states is required, cf. the beginning of Section II. For example, 4-D tomographic planning data commonly provides a limited number of different respiration states due to radiation exposure with CT or binned reconstruction and acquisition time with MRI.

The raw surfaces consist of approximately 350k 3-D points and were cropped to a region of interest covering the torso, subsequently decimated using a quadric edge collapse strategy and finally smoothed using Laplacian smoothing. Again, the motivation here is multi-modality which requires the employed algorithms to be robust w. r. t. different spatial sampling, mesh topology and partial matching issues. The final surface meshes consist of about 10k 3-D points to form the training point sets $\mathcal{P}^s$, see Fig. 5a for an illustration.

For computing the displacement fields $\mathcal{U}^s$ to match different respiration states, we represented all surfaces as distance transforms and used a volumetric non-rigid registration scheme based on level-set motion [31]. The state of fully exhale was chosen as reference $\mathcal{P}^0$. Fig. 5b exemplarily depicts a training point set along with its displacement field.

We built the motion models with an intrinsic dimensionality of $L = 3$ such that $\geq 99\%$ of the variance is explained. The resulting cumulative normalized variances are listed in Table IV in Appendix B. The importance of the third modes is considerably lower than the first two modes and we retain them solely for a consistent number of modes for all subjects. For respiratory motion analysis, we focus on the two leading modes that are depicted in Figs. 2 and 6.

For testing, the subjects were asked to perform one cycle of thoracic, abdominal and two cycles of regular breathing. We then analyzed the sequences at the full frequency of ~ 30 Hz. The corresponding number of frames are listed in Table IV in Appendix B. In contrast to the training phase, neither cropping nor mesh decimation nor Laplacian smoothing was applied.

2) Results: For model registration using (10) we set $w = 0.99$ and use a 5×5 PDA neighborhood $\mathcal{S}_{\boldsymbol{x}_n^{\boldsymbol{b}}}$, i. e. each model point $\boldsymbol{x}_n^{\boldsymbol{b}}$ is connected to its 25 nearest neighbors $\boldsymbol{y}_{n,m} \in \mathcal{S}_{\boldsymbol{x}_n^{\boldsymbol{b}}}$.

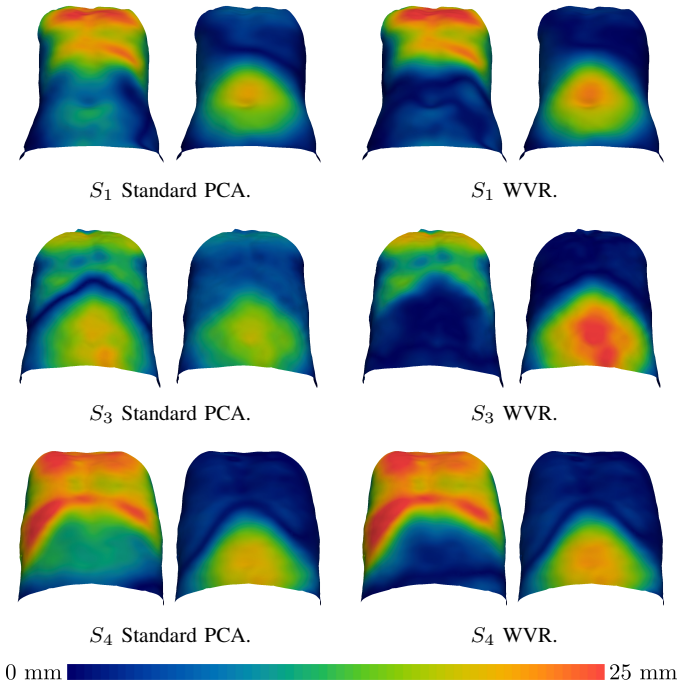


Fig. 6. The first two modes of variation for subjects S_1, S_3 and S_4, for S_2 see Fig. 2. The magnitude of the individual respiratory motion displacements in [mm] is color coded. Note that the proposed WVR scheme allows for an classification of local thoracic and abdominal motion wheres conventional PCA modes exhibit a global nature.

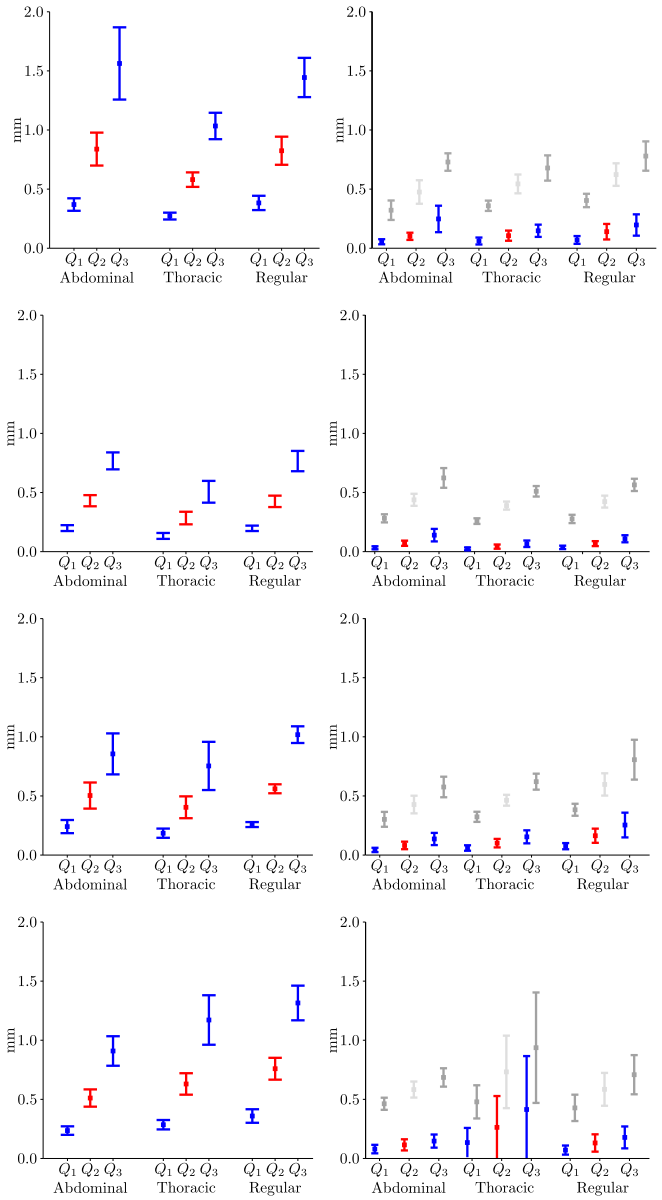
The initial guess for the model parameter was set to $\boldsymbol{b} = \boldsymbol{0}$. We assume convergence of the iterative optimization scheme when there are no considerable changes in the cost function, i. e. $|1 - |\mathcal{J}^{\text{CPD}}_{\boldsymbol{b},\sigma} / \mathcal{J}^{\text{CPD}}_{\boldsymbol{b}',\sigma'}|| < \epsilon$ where a prime denotes the estimates from the previous iteration. We set $\epsilon = 10^{-2}$.

a) Adaptability: We first assess our motion model's ability to adapt to unseen respiration states. We quantify this by the *point-to-surface* distance of registered motion models to RI surface data using $\mathcal{E}_n^{\text{M2S}} = \min\left\{\|\boldsymbol{x}_n^{\boldsymbol{b}} - \boldsymbol{y}\|_2,\ \boldsymbol{y} \in \mathcal{S}_{\boldsymbol{x}_n^{\boldsymbol{b}}}\right\}$, cf. Fig. 4. This is an unbiased metric as it does not reflect the model registration problem from (10) that relies on a *point-to-tangent* metric.

We compute the first (Q_1), second (Q_2) and third (Q_3) quartiles of $\mathcal{E}_n^{\text{M2S}}$ across all model points $\boldsymbol{x}_n^{\boldsymbol{b}}$. These quartiles are then averaged across the individual breathing sequences, see Fig. 7a. For all subjects and respiration sequences the median (Q_2) distance is always below 1.0 mm. This is in the scale of the noise level of the used RI sensor for an acquisition distance of ~ 1 m [32].

b) Robustness: To assess the robustness of our framework in the presence of noise and outliers we corrupt the RI data as $\mathcal{R}'(\boldsymbol{i}) = \mathcal{R}(\boldsymbol{i}) + \Delta r_{\boldsymbol{i}}$ prior to computing the corresponding surface representation $\mathcal{S}'(\boldsymbol{i})$. The pixel-wise offset $\Delta r_{\boldsymbol{i}}$ is drawn from a standard normal distribution to simulate noise. For 25% of the offsets which correspond to values of $|\Delta r_{\boldsymbol{i}}| > 1.15$ mm according to the standard normal distribution, an arbitrarily selected multiplier using a factor of 5 is applied to simulate outliers.

The hypothesis now is that a robust estimator is less sensitive to corruptions. We quantify this using the point-



(a) Model-to-surface distance. (b) Model-to-model difference.

Fig. 7. Model-to-surface distance and model differences in the presence of outliers with robust (colored) and conventional least-squares (gray) estimators for S_1-S_4 (top to bottom). Shown are the first to third quartiles Q_1-Q_3 averaged across all frames for abdominal, thoracic and regular breathing.

wise *model-to-model* difference $\mathcal{E}_n^{\text{M2M}} = \|\boldsymbol{x}_n^{\boldsymbol{b}} - \boldsymbol{x}_n^{\boldsymbol{b}'}\|_2$, where $\boldsymbol{x}_n^{\boldsymbol{b}} \in \mathcal{P}^{\boldsymbol{b}}$ are the points of the model $\mathcal{P}^{\boldsymbol{b}}$ registered to the original RI surface data $\mathcal{S}$ and $\boldsymbol{x}_n^{\boldsymbol{b}'} \in \mathcal{P}^{\boldsymbol{b}'}$ denotes a point of the model $\mathcal{P}^{\boldsymbol{b}'}$ aligned to the corrupted data $\mathcal{S}'$.

The quartile differences $\mathcal{E}_n^{\text{M2M}}$ across the entire surface averaged over the individual breathing sequences are depicted in Fig. 7b for the conventional least-squares estimator from (4) and the proposed robust CPD approximation from Section II-C1. The difference compared to the ground truth estimation is substantially smaller with the robust approach and, except for subject S_4 with thoracic breathing, the median difference never exceeds 0.2 mm.

c) Respiration Surrogates: For assessing our framework to deliver respiration surrogates corresponding to distinct anatomic motion we use a conventional surface monitoring technique to generate a baseline signal $\boldsymbol{s} = (s_{\text{A}}, s_{\text{T}}) \in \mathbb{R}^2$ encoding abdominal (A) and thoracic (T) surface motion. For this purpose, we manually selected two circular regions $\Omega_j \in \Omega$ with $j \in \{\text{A,T}\}$ and a diameter of approximately 5 cm for each of the subjects. The regions were placed at positions where we expect the maximum respiration amplitude w. r. t. to thoracic and abdominal motion, respectively. Similar to [15], [16], the respiratory motion signals s_j are computed as the average distance of 3-D points to the camera center according to $s_j = |\Omega_j|^{-1} \sum_{\boldsymbol{i} \in \Omega_j} \|\mathcal{S}(\boldsymbol{i})\|_2$.

TABLE I

PEARSON CORRELATION COEFFICIENTS BETWEEN CONVENTIONAL AND OUR MODEL-BASED SURROGATES FOR SEQUENCES OF ABDOMINAL, THORACIC AND REGULAR BREATHING PATTERNS.

		Thoracic		Abdominal		Regular	
		$\rho_{\text{T}}^{\text{PCC}}$	$\rho_{\text{A}}^{\text{PCC}}$	$\rho_{\text{T}}^{\text{PCC}}$	$\rho_{\text{A}}^{\text{PCC}}$	$\rho_{\text{T}}^{\text{PCC}}$	$\rho_{\text{A}}^{\text{PCC}}$
S_1	PCA	1.00	0.99	1.00	1.00	0.97	0.99
	WVR	1.00	1.00	1.00	1.00	0.98	1.00
S_2	PCA	1.00	0.31	0.86	1.00	0.96	0.99
	WVR	1.00	0.97	0.91	1.00	0.99	1.00
S_3	PCA	0.98	0.13	0.10	0.98	0.96	0.79
	WVR	0.98	0.95	0.91	1.00	0.97	0.98
S_4	PCA	0.98	1.00	1.00	0.98	0.99	0.94
	WVR	0.99	1.00	1.00	0.99	0.99	0.99

The subscripts $j \in \{\text{A,T}\}$ denote the correlation between the conventional surrogates s_j and our model-based surrogates σ_j for the abdominal (A) or thoracic (T) mode, respectively. The evaluation further opposes the standard PCA vs. the proposed WVR approach. All coefficients are statistical significant ($p_j^{\text{PCC}} \leq 0.01$).

The Pearson correlation coefficients (PCC) ρ_j^{PCC} between the baseline signal s_j and the model based surrogates σ_j are given in Table I. For the WVR model we achieve a PCC not less than 0.97. In contrast, the conventional PCA model often fails to differentiate between distinct respiration patterns with PCCs down to $\rho_{\text{T}}^{\text{PCC}} = 0.10$ as with S_3. This is due to the global nature of the conventional PCA modes. The global deformations induced by a certain mode often must be corrected by a complementary mode in order to fit the surface data. This yields a score on the model parameter b_j governing the surrogate σ_j though there is no anatomical equivalent. Due to its sparsity, the WVR model does not suffer from this problem. Over all sequences, subjects and surrogates we found a mean PCC of 0.91 for the PCA and 0.98 for the WVR model.

d) Run-time Evaluation: Our framework is designed for real-time computation on modern GPU architectures. Besides the real-time capability of multi-view RI reconstruction in Section II-B that we demonstrated in preceding work [20], the approximated CPD framework outlined in Section II-C1 can be computed efficiently on GPU architectures. We implemented the framework using the CUDA programming model. The evaluation is performed on a PC equipped with an Intel Core i7 3770K CPU and an NVIDIA GTX 680 GPU.

The run-times for the different steps of our framework are given in Table II. For multi-view RI data fusion and surface reconstruction from Section II-B with subsequent post-processing we achieve steady run-times of ~ 5.5 ms. In contrast, the run-times for the motion model registration

TABLE II
RUN-TIME EVALUATION OF THE PROPOSED FRAMEWORK.

	Fusion [ms]	CPD P2T		CPD P2P	
		Iterations	Run-time [ms]	Iterations	Run-time [ms]
S_1	5.1 ± 0.4	15.8 ± 5.5	18.5 ± 5.5	29.1 ± 3.6	28.9 ± 3.5
S_2	5.4 ± 0.5	9.67 ± 1.4	12.2 ± 1.8	34.0 ± 5.3	33.9 ± 4.9
S_3	5.6 ± 0.5	12.1 ± 3.1	15.6 ± 3.6	32.2 ± 8.1	33.0 ± 7.6
S_4	5.5 ± 0.5	14.6 ± 6.1	16.2 ± 5.8	37.8 ± 6.2	37.2 ± 5.8

Evaluated is the multi-view RI data fusion with surface reconstruction (cf. Section II-B) and the model registration using our point-to-tangent (P2T) opposed to conventional point-to-point (P2P) approximated CPD schemes (cf. Section II-C).

show more variations due to the fact that more iterations are required for instantaneous respiration states that are considerably different from the initial guess. For our point-to-tangent metric (CPD P2T), we require 13 iterations in average which corresponds to ~ 16 ms. For comparison, the conventional point-to-point metric (CPD P2P) requires 33 iterations and ~ 33 ms. In conclusion, our framework (CPD P2T) achieves run-times of about 20 ms or 50 Hz.

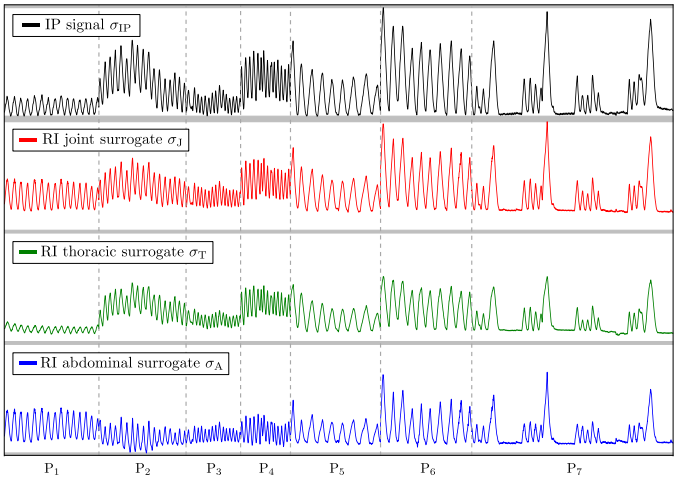


Fig. 8. Qualitative visualization of the IP signal σ_{IP} compared to the joint (σ_{J}), thoracic (σ_{T}) and abdominal (σ_{A}) model-based RI surrogates for subject S_5 and different respiration patterns. The signals divide into abdominal (P_1), thoracic (P_2), fast shallow (P_3), fast strong (P_4), slow shallow (P_5), slow strong (P_6) and breath hold (P_7) breathing instructions. For visualization, the IP signal is scaled to $[\min(\sigma_{\text{IP}}), \max(\sigma_{\text{IP}})]$ and the RI surrogates to $[\min(\sigma_{\text{J}}, \sigma_{\text{T}}, \sigma_{\text{A}}), \max(\sigma_{\text{J}}, \sigma_{\text{T}}, \sigma_{\text{A}})]$.

B. Comparison to Established Respiration Sensor

This experiment is concerned with a comparison of the respiratory motion surrogates obtained from our WVR model and an established baseline respiration sensor that does not rely on body surface deformations but instead directly measures the change of lung volume.

1) Setup and Data: For this experiment data from three additional subjects S_5–S_7 were collected. As baseline, we use a PMM2 sensor (Siemens Healthcare, Erlangen, Germany) which uses the measurement principle of impedance pneumography (IP). The IP sensor sends high frequency current through bio-potential electrodes that are attached to the left and right side of the thorax and the side of the belly. Based on the change of impedance caused by the change of lung volume, a one-dimensional respiration signal was obtained at 250 Hz.

Simultaneously, RI surface data was acquired with two Asus Xtion PRO RI sensors with a frame-rate of ~ 30 Hz and a resolution of $\Omega_{1,2} = \mathbb{R}^{320\times240}$. The same multi-view fusion, pre-processing, model generation and registration methods as in Section III-A were used. To match the RI frame-rate, the IP signal was uniformly downsampled to 30 Hz.

Similar to the experiments in Section III-A, the subjects were asked to perform free abdominal and thoracic breathing to train our 4-D motion model, see Table V in Appendix B for a detailed listing of the number of training samples and properties of the proposed WVR model.

For testing we use a protocol that specifies predefined breathing instructions over a period of about 7 minutes which resulted in approximately 12k respiration measurements. First, abdominal and thoracic breathing patterns were to be performed with a duration of about 60 s, each. Subsequently, the subjects were instructed to breath at fixed frequencies of 0.5 and 0.17 Hz. These phases subdivide into shallow and strong breathing where the duration was set to approximately 30 s for fast, and about 60 s for slow breathing. Finally, breath-holds with a duration of 15 s, each, were to be performed over a period of approximately 120 s. The resulting number of frames that were used for evaluation are detailed in Table V in Appendix B.

Though being acquired simultaneously, the RI surrogates and the IP signal exhibited a temporal shift caused by bus and driver latencies. To bring both signals into congruence, we phase-wise estimated the shift that maximizes the correlation coefficient between the shifted IP signal and the RI surrogate $\sigma_{\text{J,A,T}}$ from (5). The rational behind this alignment scheme is that the temporal shift is not uniform across the entire evaluation period and that at least one RI surrogate is supposed to explain the IP signal within a localized window.

2) Results: Fig. 8 exemplarily depicts the registered IP signal and the proposed model-based RI surrogates for one of the subjects over the entire evaluation period. Noticeable are the distinct breathing instructions and varying degrees of correlation between the IP signal and the model-based surrogates. This is most distinct for phase P_2 where the abdominal surrogate σ_{A} contradicts the IP sensor regarding both signal scale and shape whereas the thoracic surrogate σ_{T} follows the IP signal.

For quantification, we first computed the Pearson correlation coefficients between the shifted IP signal and our respiration surrogates for the individual sequences P_i. The results are listed in Table III. For the joint signal σ_{J}, an average PCC of 0.96 ± 0.04 with $\rho_{\text{J}}^{\text{PCC}} \geq 0.97$ for 19 out of 21 evaluations and a minimum of 0.92 was found across the individual phases and subjects. In contrast, the thoracic and abdominal surrogates show substantial variations regarding the correlation with the IP signal and an average PCC of 0.79 ± 0.35 and 0.84 ± 0.28 for the abdominal and thoracic surrogate, respectively. Here, the most prominent results are for subject S_6 and phase P_1 where there is no correlation in the thoracic

TABLE III
PEARSON CORRELATION COEFFICIENTS BETWEEN THE SHIFTED IP SENSOR SIGNAL AND OUR WVR SURROGATES FOR PHASES OF DIFFERENT BREATHING PATTERNS AND ACROSS ALL RESPIRATION SAMPLES.

	All phases			P_1			P_2			P_3			P_4			P_5			P_6			P_7		
	ρ_J^{PCC}	ρ_T^{PCC}	ρ_A^{PCC}	ρ_J^{PCC}	ρ_T^{PCC}	ρ_A^{PCC}	ρ_J^{PCC}	ρ_T^{PCC}	ρ_A^{PCC}	ρ_J^{PCC}	ρ_T^{PCC}	ρ_A^{PCC}	ρ_J^{PCC}	ρ_T^{PCC}	ρ_A^{PCC}	ρ_J^{PCC}	ρ_T^{PCC}	ρ_A^{PCC}	ρ_J^{PCC}	ρ_T^{PCC}	ρ_A^{PCC}	ρ_J^{PCC}	ρ_T^{PCC}	ρ_A^{PCC}
S_5	0.96	0.97	0.73	0.98	0.85	**0.98**	0.94	**0.96**	0.63	0.98	**0.98**	0.92	0.97	**0.99**	0.80	**0.99**	0.98	0.94	0.99	**0.99**	0.96	**0.99**	0.98	0.98
S_6	0.93	0.93	0.90	**0.97**	−0.04	0.97	**0.98**	0.96	0.97	**0.98**	0.55	0.96	**0.99**	0.93	0.97	**0.99**	0.93	0.98	**1.00**	0.99	0.99	**0.98**	0.88	0.96
S_7	0.96	0.99	0.16	0.92	**0.99**	0.86	0.97	**0.99**	−0.47	0.97	**0.97**	0.84	0.98	**0.98**	0.09	**0.98**	0.97	0.96	0.99	**1.00**	0.63	0.98	**0.98**	0.74

The subscripts $j \in \{J,A,T\}$ denote the correlation of the IP signal w. r. t. the joint (σ_J), abdominal (σ_A) and thoracic (σ_T) WVR model surrogates. The evaluation phases divide into abdominal (P_1), thoracic (P_2), fast shallow (P_3), fast strong (P_4), slow shallow (P_5), slow strong (P_6) and breath hold (P_7) breathing instructions, cf. Fig. 8. The surrogate yielding the best correlation to estimate the signal shift is highlighted.

mode ($\rho_T^{PCC} = -0.04$) but a very strong correlation in the abdominal part ($\rho_A^{PCC} = 0.97$).

These results show that an RI based respiratory motion analysis regarding change of lung volume requires surrogates based on distinct spatially localized surface displacements as well as their superposition. However, which surrogate best correlates to the IP signal depends on the breathing pattern and the subject. This is no flaw in our model-based respiration analysis formulation but instead indicates that the 1-D IP signal as a measure of lung volume change is not capable to describe the entire spectrum of surface deformations.

As an overall performance metric, we computed the PCCs across the entirety of available breathing samples. This avoids any bias that might occur in the phase-wise correlation. Here, we found PCCs of 0.93–0.96 for the joint, 0.93–0.99 for the thoracic, and 0.16–0.90 for the abdominal surrogates. Compared to the phase-wise evaluation, the performance of the joint signal slightly decreased whereas the thoracic surrogate shows a higher correlation w. r. t. the IP signal. This effect is solely due to the mathematical foundations of the PCC.

IV. CONCLUSION AND FUTURE WORK

We proposed a real-time respiratory motion analysis framework that uses pre-procedurally trained 4-D shape priors to drive intra-procedural RI-based respiration analysis. As the major contribution, we presented a sparse motion model that allows for an unsupervised differentiation of thoracic and abdominal breathing. For intra-procedural respiration analysis, we presented a robust and fast method to register our motion model to body surface data computed from multi-view RI.

Our model-based formulation allows to derive anatomically plausible low-dimensional respiration surrogates encoding surface extents for different body regions. Compared to existing RI-based respiration analysis approaches, our method is not intrusive, does not rely on heuristic surface partitioning and does not require manual interaction.

One key finding of our work is that motion models can be constructed using PCA with a small number of variation modes. We showed that respiration induced surface deformations can be reconstructed by model-based dense abdominal and thoracic displacement fields with a median error < 1.0 mm. This indicates that such body surface deformations are governed by only two factors and that displacements in different thoracic (or abdominal) regions can be inferred from each other. In this regard, we found a statistically significant Pearson correlation coefficient of 0.98 between our model-based surrogates and conventional surrogates encoding thoracic and abdominal movement.

Regarding practical aspects, we demonstrated the robustness and efficiency of an approximative model registration scheme extending the *Coherent Point Drift* by a point-to-tangent metric. Compared to conventional point-to-point metrics we require half of the number of iterations resulting in run-times of ~ 50 Hz for a GPU-based implementation.

Compared to impedance pneumography as an intrusive and fundamentally different respiration measurement principle, our model-based surrogates yield a PCC ≥ 0.93 across seven different breathing phases covering a period of about 7 minutes with $\sim$ 12k respiration samples. For the individual phases, we showed the varying degree of correlation of surrogates based on abdominal, thoracic or superimposed body surface deformations.

Especially the results from the last experiment raise the question whether external respiration surrogates or direct internal signals are more suited for respiration analysis. This is of particular importance when the surrogates or signals are used to predict internal target movement. However, this aspect is not within the scope of this work and must be addressed in future research that also must be concerned with the evaluation of our proposed respiratory motion analysis framework in a clinical setting. This includes both the evaluation on non-healthy subjects as well as the integration into clinical workflows.

ACKNOWLEDGMENT

We thank Metrilus GmbH, Erlangen, Germany, for providing the camera equipment and the calibration framework used in the IP experiment of this work.

APPENDIX A
MULTI-LINK MODEL-TO-TANGENT DISTANCE

As both the posteriors $p(\boldsymbol{x}_n^{\boldsymbol{b}}|\boldsymbol{y}_m)$ from (11) and the kernel width σ are constants when optimizing (10) w. r. t. $\boldsymbol{b}$, we define $\alpha_{n,m} = \sigma^{-2} p(\boldsymbol{x}_n^{\boldsymbol{b}}|\boldsymbol{y}_m)$. We expand (10) using the modes of variation $\boldsymbol{e}_l$ column-wise contained in $\boldsymbol{P} = (p_{i,j})$ and $\boldsymbol{b} = (b_1, \ldots, b_L)^\top$ point-wise as:

$$\mathcal{J}_{\boldsymbol{b}}^{\text{CPD}} = \sum_{n=1}^{N} \sum_{m=1}^{|\mathcal{S}_n|} \alpha_{n,m} \left(\left(\sum_{l=1}^{L} b_l \boldsymbol{e}_{l,n} - \boldsymbol{y}_{n,m} \right) \cdot \boldsymbol{n}_{n,m} \right)^2 ,$$

where $\boldsymbol{e}_{l,n} = (p_{3\cdot(n-1)+1,l}, p_{3\cdot(n-1)+2,l}, p_{3\cdot(n-1)+3,l})^\top$ and $\boldsymbol{y}_{n,m} \in \mathcal{S}_{\boldsymbol{x}_n^{\boldsymbol{b}}}$ are the nearest neighbors for the n-th point in our approximated CPD framework, cf. Section II-C2. Computing the partial derivative of $\mathcal{J}_{\boldsymbol{b}}$ w. r. t. one specific motion model parameter b_k and equating to zero yields:

$$\sum_{n=1}^{N}\sum_{m=1}^{|\mathcal{S}_n|}\sum_{l=1}^{L}\alpha_{n,m}b_l\eta_{l,n,m}\eta_{k,n,m} \stackrel{!}{=} \sum_{n=1}^{N}\sum_{m=1}^{|\mathcal{S}_n|}\alpha_{n,m}\zeta_m\eta_{k,n,m},$$

where $\eta_{l,n,m} = \boldsymbol{e}_{l,n}\cdot\boldsymbol{n}_{n,m}$ and $\zeta_m = \boldsymbol{y}_{n,m}\cdot\boldsymbol{n}_{n,m}$. For the vector $\boldsymbol{b}$ this can be written using matrix-vector notation as:

$$\sum_{n=1}^{N}\sum_{m=1}^{|\mathcal{S}_n|}\alpha_{n,m}\begin{pmatrix}\eta_{1,n,m}\eta_{1,n,m} & \cdots & \eta_{1,n,m}\eta_{L,n,m}\\ \vdots & \ddots & \vdots\\ \eta_{L,n,m}\eta_{1,n,m} & \cdots & \eta_{L,n,m}\eta_{L,n,m}\end{pmatrix}\boldsymbol{b} = \sum_{n=1}^{N}\sum_{m=1}^{|\mathcal{S}_n|}\alpha_{n,m}\begin{pmatrix}\zeta_m\eta_{1,n,m}\\ \vdots\\ \zeta_m\eta_{L,n,m}\end{pmatrix}. \quad (12)$$

This is a linear equation system in the form $\boldsymbol{A}\boldsymbol{b} = \boldsymbol{c}$ where the symmetric matrix $\boldsymbol{A} \in \mathbb{R}^{L\times L}$ is positive definite for weights and diagonal elements > 0. By definition, $\alpha_{n,m} > 0$, cf. (11). Further, $\sum_{n=1}^{N}\sum_{m=1}^{|\mathcal{S}_n|}\eta_{l,n,m}^2 \geq 0$. These scalars are zero iff *all* motion model directions $\boldsymbol{e}_{l,n}$ are perpendicular to the corresponding normals $\boldsymbol{n}_{n,m}$. Thus, *all* points would move parallel to the surface which is an invalid degenerated model.

Regarding real-time aspects, we note that (12) essentially breaks down to the summation of matrices. This problem can be calculated very efficiently on GPU architectures using tree-based parallel reduction techniques [33].

For efficient nearest neighbor computation, we represent the RI surface $\mathcal{S}$ as a multi-channel 2-D texture where the individual channels hold the x, y and z coordinates, respectively. For a point $\boldsymbol{x}_n^{\boldsymbol{b}}$ the closest points $\boldsymbol{y}_{n,m} \in \mathcal{S}_{\boldsymbol{x}_n^{\boldsymbol{b}}}$ are then obtained by texture look-ups w. r. t. the index neighborhood $\Omega_{\boldsymbol{x}_n^{\boldsymbol{b}}}$, cf. Fig. 4. This scheme exploits the cache layout of GPU textures being designed for spatial locality in an index-based sense.

APPENDIX B
DETAILED DATA SPECIFICATION

TABLE IV
DATA FOR METHODOLOGY ASSESSMENT IN SECTION III-A.

	Training					Testing		
	|T|	|A|	λ'_1	λ'_2	λ'_3	|T|	|A|	|R|
S_1	6	5	0.76 (0.79)	0.98 (0.98)	0.99 (0.99)	167	112	272
S_2	6	6	0.72 (0.76)	0.99 (0.99)	0.99 (0.99)	118	122	324
S_3	5	5	0.53 (0.67)	0.98 (0.98)	0.99 (0.99)	235	156	354
S_4	7	6	0.84 (0.85)	0.96 (0.96)	0.99 (0.99)	89	168	328

Listed are the number of frames in thoracic (T), abdominal (A) and regular (R) breathing sequences. For the training stage, the normalized cumulative variances $\lambda'_i = \sum_{s=1}^{i}\lambda_s / \sum_{s=1}^{3}\lambda_s$ are reported. Plain values denote the proposed WVR model and bracketed values denote conventional PCA.

TABLE V
DATA FOR THE IP SENSOR EXPERIMENT IN SECTION III-B.

	Training					Testing ($\lvert P_i\rvert \cdot 10^3$)						
	|T|	|A|	λ'_1	λ'_2	λ'_3	$\lvert P_1\rvert$	$\lvert P_2\rvert$	$\lvert P_3\rvert$	$\lvert P_4\rvert$	$\lvert P_5\rvert$	$\lvert P_6\rvert$	$\lvert P_7\rvert$
S_5	7	6	0.64	0.93	0.97	1.6	1.5	0.9	0.9	1.5	1.6	3.4
S_6	7	6	0.87	0.97	0.99	1.6	1.6	0.8	0.8	1.5	1.7	3.4
S_7	7	6	0.50	0.94	0.97	1.5	1.6	0.9	0.8	1.6	1.5	3.3

Shown are the number of frames in the thoracic (T) and abdominal (A) training sequences and the resulting WVR model variances λ'_i (cf. Table IV). For the testing stage, the number of frames in phases with predefined respiration patterns P_1–P_7 are listed (cf. Section III-B1 and Fig. 8).

REFERENCES

[1] C. Bert and M. Durante, "Motion in radiotherapy: particle therapy," *Physics in Medicine and Biology*, vol. 56, no. 16, pp. R113–44, 2011.

[2] S. S. Korreman, "Motion in radiotherapy: photon therapy," *Physics in Medicine and Biology*, vol. 57, no. 23, pp. R161–91, 2012.

[3] H. Hof *et al.*, "Stereotactic single-dose radiotherapy of stage i nonsmall-cell lung cancer (nsclc)," *International Journal of Radiation Oncology Biology Physics*, vol. 56, no. 2, pp. 335 – 341, 2003.

[4] P. J. Keall *et al.*, "The management of respiratory motion in radiation oncology report of aapm task group 76a," *Medical Physics*, vol. 33, no. 10, pp. 3874–3900, 2006.

[5] D. Verellen *et al.*, "Gating and tracking, 4d in thoracic tumours," *Cancer/Radiothérapie*, vol. 14, no. 67, pp. 446 – 454, 2010.

[6] T. Willoughby *et al.*, "Quality assurance for nonradiographic radiotherapy localization and positioning systems: Report of task group 147," *Medical Physics*, vol. 39, no. 4, pp. 1728–1747, 2012.

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

[8] A. Fassi *et al.*, "Tumor tracking method based on a deformable 4d CT breathing motion model driven by an external surface surrogate," *International Journal of Radiation Oncology*Biology*Physics*, vol. 88, no. 1, pp. 182 – 188, 2014.

[9] S. Bauer *et al.*, "Marker-less Reconstruction of Dense 4-D Surface Motion Fields Using Active Laser Triangulation for Respiratory Motion Management," in *Proc MICCAI 2012, Part I*, vol. LNCS 7510, 2012, pp. 414–421.

[10] G. J. Price *et al.*, "Real-time optical measurement of the dynamic body surface for use in guided radiotherapy," *Physics in Medicine and Biology*, vol. 57, no. 2, pp. 415–36, 2012.

[11] B. L. Lindl *et al.*, "Topos: A new topometric patient positioning and tracking system for radiation therapy based on structured white light," *Medical Physics*, vol. 40, no. 4, p. 042701, 2013.

[12] C. Schaller, J. Penne, and J. Hornegger, "Time-of-Flight Sensor for Respiratory Motion Gating," *Medical Physics*, vol. 35, no. 7, pp. 3090–3093, 2008.

[13] H. Fayad *et al.*, "A patient specific respiratory model based on 4D CT data and a time of flight camera (TOF)," in *Proc IEEE NSS/MIC*, 2009, pp. 2594–2598.

[14] S. Hughes *et al.*, "Assessment of two novel ventilatory surrogates for use in the delivery of gated/tracked radiotherapy for non-small cell lung cancer," *Radiotherapy and Oncology*, vol. 91, no. 3, pp. 336 – 341, 2009.

[15] H. Fayad *et al.*, "Technical note: Correlation of respiratory motion between external patient surface and internal anatomical landmarks," *Medical Physics*, vol. 38, no. 6, pp. 3157–3164, 2011.

[16] J. Xia and R. A. Siochi, "A real-time respiratory motion monitoring system using kinect: Proof of concept," *Medical Physics*, vol. 39, no. 5, pp. 2682–2685, 2012.

[17] J. Schaerer *et al.*, "Multi-dimensional respiratory motion tracking from markerless optical surface imaging based on deformable mesh registration," *Physics in Medicine and Biology*, vol. 57, no. 2, pp. 357–373, 2012.

[18] D. Spinczyk, A. Karwan, and M. Copik, "Methods for abdominal respiratory motion tracking," *Computer Aided Surgery*, vol. 19, no. 1-3, pp. 34–47, 2014.

[19] J. Wasza, S. Bauer, and J. Hornegger, "Real-time Motion Compensated Patient Positioning and Non-rigid Deformation Estimation Using 4-D Shape Priors," in *Proc MICCAI 2012, Part II*, vol. LNCS 7511, 2012, pp. 576–583.

[20] J. Wasza, S. Bauer, and J. Hornegger, "Real-Time Respiratory Motion Analysis Using Manifold Ray Casting of Volumetrically Fused Multi-view Range Imaging," in *Proc MICCAI 2013, Part II*, vol. LNCS 8150, 2013, pp. 116–123.

[21] A. Myronenko and X. Song, "Point set registration: Coherent point drift," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 12, pp. 2262 –2275, 2010.

[22] O. Taubmann *et al.*, "Prediction of Respiration-Induced Internal 3-D Deformation Fields From Dense External 3-D Surface Motion," in *Proceedings of the 28th International Congress and Exhibition*, ser. International Journal of Computer Assisted Radiology and Surgery, H. U. Lemke *et al.*, Eds., Heidelberg, 2014, pp. 33–34.

[23] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification (2nd Edition)*. Wiley-Interscience, 2000.

[24] H. Kaiser, "The varimax criterion for analytic rotation in factor analysis," *Psychometrika*, vol. 23, no. 3, pp. 187–200, 1958.

[25] J. Wasza *et al.*, "Sparse Principal Axes Statistical Surface Deformation Models for Respiration Analysis and Classication," in *Bildverarbeitung für die Medizin (BVM) 2012*, 2012, pp. 316–321.

[26] M. B. Stegmann, K. Sjöstrand, and R. Larsen, "Sparse Modeling of Landmark and Texture Variability using the Orthomax Criterion," in *Proc SPIE*, vol. 6144, 2006, pp. 61 441G1–61 441G.12.

[27] B. Curless and M. Levoy, "A volumetric method for building complex models from range images," in *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*, ser. SIGGRAPH '96. ACM, 1996, pp. 303–312.

[28] S. Placht *et al.*, "Fast time-of-flight camera based surface registration for radiotherapy patient positioning," *Medical Physics*, vol. 39, no. 1, pp. 4–17, 2012.

[29] S. Rusinkiewicz and M. Levoy, "Efficient variants of the icp algorithm," in *Proc. International Conference on 3-D Digital Imaging and Modeling, 2001*, 2001, pp. 145–152.

[30] J. Wasza, S. Bauer, and J. Hornegger, "Real-time Preprocessing for Dense 3-D Range Imaging on the GPU: Defect Interpolation, Bilateral Temporal Averaging and Guided Filtering," in *IEEE International Conference on Computer Vision (ICCV) Workshops*, 2011, pp. 1221–1227.

[31] B. Vemuri *et al.*, "Image registration via level-set motion: Applications to atlas-based segmentation," *Medical Image Analysis*, vol. 7, no. 1, pp. 1 – 20, 2003.

[32] J. Smisek, M. Jancosek, and T. Pajdla, "3d with kinect," in *IEEE International Conference on Computer Vision (ICCV) Workshops*, 2011, pp. 1154–1160.

[33] M. Harris, S. Sengupta, and J. D. Owens, "Parallel Prefix Sum (Scan) with CUDA," in *GPU Gems 3*. Addison Wesley, Aug 2007.