Multi-modal Surface Registration for Markerless Initial Patient Setup in Radiation Therapy using Microsoft's Kinect Sensor

Sebastian Bauer[1], Jakob Wasza[1], Sven Haase[1], Natalia Marosi[3], 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
[3]Siemens AG, Healthcare Sector, Erlangen, Germany

sebastian.bauer@cs.fau.de

Abstract

In radiation therapy, prior to each treatment fraction, the patient must be aligned to computed tomography (CT) data. Patient setup verification systems based on range imaging (RI) can accurately verify the patient position and adjust the treatment table at a fine scale, but require an initial manual setup using lasers and skin markers. We propose a novel markerless solution that enables a fully-automatic initial coarse patient setup. The table transformation that brings template and reference data in congruence is estimated from point correspondences based on matching local surface descriptors. Inherently, this point-based registration approach is capable of coping with gross initial misalignments and partial matching. Facing the challenge of multi-modal surface registration (RI/CT), we have adapted state-of-the-art descriptors to achieve invariance to mesh resolution and robustness to variations in topology. In a case study on real data from a low-cost RI device (Microsoft Kinect), the performance of different descriptors is evaluated on anthropomorphic phantoms. Furthermore, we have investigated the system's resilience to deformations for mono-modal RI/RI registration of data from healthy volunteers. Under gross initial misalignments, our method resulted in an average angular error of 1.5° and an average translational error of 13.4 mm in RI/CT registration. This coarse patient setup provides a feasible initialization for subsequent refinement with verification systems.

1. Introduction

Patient setup reproducibility is one of the major challenges in fractionated radiation therapy (RT) planning and treatment. Precise alignment is a mandatory prerequisite for the success of RT, improving the balance between complications and treatment and providing the fundamental basis for high-dose and small-margin irradiation application. Prior to each fraction, the patient must be accurately aligned w.r.t. the target isocenter that has been localized in planning CT data. For setup verification and correction, radiographic portal imaging, cone-beam CT and CT-on-rails may be applied. However, this involves additional radiation exposure to the patient. Non-radiographic techniques that locate electromagnetic fiducials [15] are an accurate alternative, but require the patient to be eligible for the invasive procedure of marker implantation.

Over the past few years, due to advances in sensor technology, several devices for non-radiographic and non-

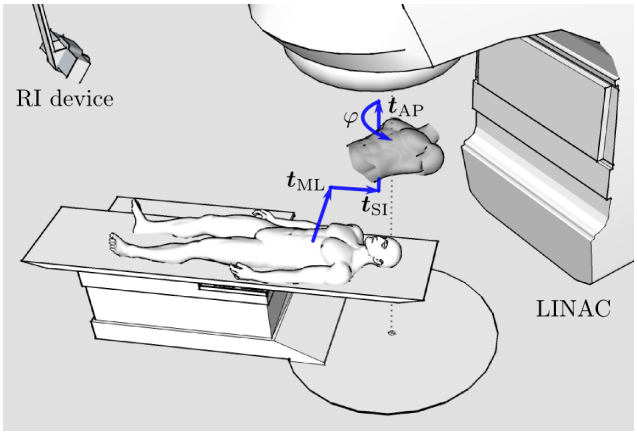


Figure 1. Schematic illustration of the proposed automatic initial patient setup: The patient's intra-fractional surface is acquired with an RI device and registered to planning CT data (depicted in gray). The estimated transformation (blue) that brings template and reference in congruence is then applied to the treatment table.

invasive patient setup and monitoring based on range imaging (RI) have been introduced in clinical practice [4, 8, 12, 18, 20].

Regardless of the particular RI technology, the systems provide a complete, metric and precise 3-D surface model of the patient. They also estimate the table transformation that brings a pre-defined region of the intra-fractional patient surface in congruence with a reference at a fine scale. A number of studies have shown that these techniques obtain a high degree of precision for patient setup for thoracic and abdominal tumor locations [2, 8, 12, 14, 20]. In addition, most RI systems are able to capture dense 3-D surface data in real-time, allowing for continuous patient monitoring during the course of a treatment session. However, existing solutions are designed with a focus on setup verification and require an initial patient alignment with conventional techniques using lasers and skin markers [2, 8, 12, 14]. This manual initial coarse setup is both a time-consuming and tedious procedure.

In this paper, we propose an RI-based approach that enables a markerless and automatic initial coarse RT patient setup, superseding the need for lasers and skin markers. Without spending extra time, the initial patient setup is performed in an unlabored manner. Since this is only an initial alignment to be followed by position refinement [1, 4, 8, 20], the accuracy requirements are rather low (isocenter position within $\pm$ 50 mm) [8]. The proposed method can be applied to reference surfaces either acquired by the RI device prior to the first fraction, or extracted from planning CT data. It estimates the optimal table transformation from point correspondences between local features. We have extended state-of-the-art descriptors to handle distinct mesh resolutions and topological variations that occur due to the low signal-to-noise ratio (SNR) of RI sensors. By design, the approach can handle gross initial misalignments and cope with partial matching [5, 10, 11], see Fig. 1. This is a fundamental prerequisite for both the mono-modal (RI/RI) and multi-modal (RI/CT) case, where the intra-fractionally covered template surface region may differ severely from the reference. Typically, the field of view of the RI device is considerably larger than the CT scan volume.

2. Related Work

Matching local invariant feature descriptors is a key component of a variety of computer vision tasks in the 2-D and 3-D domain such as registration, object recognition, scene reconstruction or similarity search in databases. 3-D surface registration is more relevant to the topic of this paper. Thus, we will focus our discussion on this subfield. In this context, a popular method for solving the alignment problem of two or more sets of points or surfaces is the iterative closest point (ICP) algorithm [3]. However, in the presence of gross misalignments, the algorithm depends on a proper initialization in order to prevent the iterative transformation estimation from being stalled by local minima [19]. Furthermore, ICP is not designed to handle partial overlap in case of occlusions, clutter and viewpoint changes.

In partial 3-D surface matching, the trend is towards methods that establish point correspondences from matching local feature descriptors. Typically, the descriptors encode the surface geometry of the underlying data in a limited support region around an interest point [6, 9, 13, 23, 25]. The correspondences can then be used to find the transformation that maximizes the alignment. For a comprehensive survey of 3-D surface descriptors see the work of Bustos et al. [5]. Among the first descriptors in the field were point signatures [6] and spin images [13]; the latter remains one of the most popular methods for 3-D surface description to date. Frome et al. extended the concept of shape contexts to the 3-D domain [9]. Recent approaches focused on hybrid models at the intersection between signatures and histograms to balance descriptiveness and robustness [23]. In the planar domain, the majority of successful descriptors such as HOG [7], SIFT [16], and RIFF [22] rely on histogram representations. In the context of local descriptors, histograms trade-off descriptive power and positional precision for robustness and repeatability by compressing geometry into bins [23], thus being an appropriate choice for noisy data. Based on an orthographic depth representation of the local 3-D surface topology in a planar patch, 2-D descriptors can be applied to surface data in a straightforward manner. Furthermore, concepts from 2-D feature description can be extended to the 3-D domain. For instance, inspired by the performance of HOG, Zaharescu et al. extended the descriptor to scalar fields defined on 2-D manifolds (MeshHOG) [25].

We propose extensions to state-of-the-art descriptors that enable its application for robust multi-modal surface registration. First, we have modified the MeshHOG descriptor to achieve invariance to mesh resolution and robustness to topological variations due to noise and quantization effects. Second, we introduce a scheme that extends the 2-D RIFF descriptor to the domain of 3-D surfaces. Our method is based on a correspondence search engine that enables partial matching, and that is resilient w.r.t. minor deformations that occur in practice due to body distortion and respiratory motion. To our knowledge, the automation of initial coarse setup in RT has not been addressed yet.

3. Methods

The proposed framework is composed of three stages (see Fig. 3). First, the local topology of the template (RI) and reference surface (RI/CT) is encoded by our modified descriptors that are specifically designed to handle surface data from distinct modalities. Second, point correspon-

dences are established by descriptor matching and pruned by incorporating a geometric consistency analysis. Third and last, the optimal rigid-body transformation is estimated w.r.t. the coordinate system of the treatment table. The initial patient setup can then be performed by adjusting the table based on the estimated transformation.

3.1. Surface Registration Framework

Our surface registration framework relies on feature descriptors that encode the local geometric topology in a translation- and rotation-invariant manner. However, the descriptors are not invariant to scale on purpose, as we incorporate the metric scale of the anatomical surface topology as an important characteristic. Placing great importance on robustness and repeatability, we have extended two alternative descriptors (MeshHOG, RIFF) for multi-modal application. Both rely on histograms of oriented gradients (HOG) [7] which have shown to be leading edge in terms of classification performance for the 2-D case. As a baseline, we compare these HOG-like descriptors to the well-established technique of spin images [13]. Below, we outline the descriptors' functional principles and inevitable adaptations for the problems at hand.

MeshHOG. In developing features for this particular multi-modal scenario, one has to focus on two attributes: robustness to topology variations and invariance to mesh resolution. Hence, the traditional MeshHOG approach had to be adapted accordingly. The descriptor may be considered as a generalization of the concept of HOG from planar image domains to non-planar 2-D manifolds. It encodes the local spatial distribution of a gradient vector field $\mathcal{F} = \{\nabla f(\boldsymbol{x})\}$ derived from a scalar function $f(\boldsymbol{x})$. Given a surface point $\boldsymbol{x}_i$ and a support region $\mathcal{N}_i$, the gradients $\nabla f(\boldsymbol{x}_j) \in \mathbb{R}^3,\ \boldsymbol{x}_j \in \mathcal{N}_i$ are projected onto the orthogonal planes of a unique and invariant local reference frame. Subsequently, an orientation histogram binning is performed

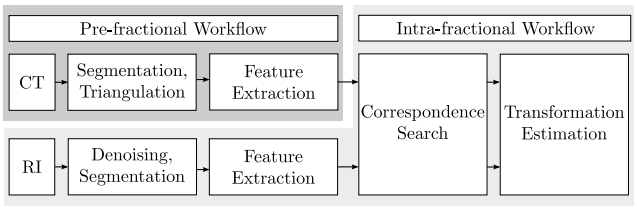
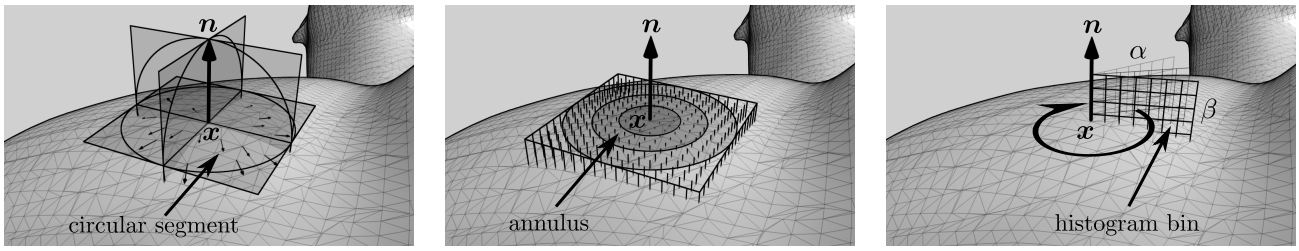


Figure 3. Flowchart of the feature-based registration framework. Data propagates from left to right. The shading indicates the parts of the workflow that are performed prior to the first fraction.

w.r.t. circular segments (Fig. 3a). In this work, the scalar function $f(\boldsymbol{x})$ characterizes the local surface geometry. In particular, in order to cope with the low SNR of RI data, we propose the signed distance of a point to the best fitting plane of its local neighborhood instead of second order derivatives such as curvature measures [25]. The incorporation of additional photometric information [25] is unfeasible for applications involving untextured CT data. Zaharescu et al. computed the gradient vectors using a discrete operator that relies on adjacent vertices, restricting the approach to uniformly sampled triangular meshes. To be able to cope with arbitrary mesh representations and resolutions in a multi-modal surface registration setup, we have replaced the original operator by a circular uniform surface sampling technique similar to [17].

RIFF. By definition, the conventional rotation-invariant fast features (RIFF) operate in the 2-D image domain [22]. We have developed a scheme that extends the RIFF concept to the domain of 3-D surfaces. As an initial step, we encode the 3-D surface topology in the neighborhood of $\boldsymbol{x}_i$ as a local orthographic depth representation w.r.t. the tangent plane defined by the normal $\boldsymbol{n}_i$. For this 2-D patch, we then compute the RIFF descriptor [22] (Fig. 3b). First, rotation invariance is achieved by performing a radial gradient transform. Second, the patch is subdivided into annular



(a) (b) (c)

Figure 2. Functional principle of the surface descriptors: The MeshHOG descriptor (a) projects a gradient vector field onto circular segments of the orthogonal planes of a local coordinate system. For the RIFF descriptor (b), the surface topology is expressed as a 2-D depth representation where rotation-invariant gradients are binned for different annuli. Spin images (c) encode cylindrical coordinates (α, β) in a 2-D histogram.

spatial bins. Third, given the rotation-invariant gradients, an orientation histogram binning (HOG) is performed. The feature descriptor represents a concatenation of histograms from different annuli.

Spin Images. For comparison purposes we also used spin images. Introduced more than a decade ago [13], spin images enjoy great popularity for surface matching. Given an oriented surface point $\boldsymbol{x}_i \in \mathbb{R}^3$ with its associated normal $\boldsymbol{n}_i$, its spin image is generated as follows: The entirety of points $\boldsymbol{x}_j$ within a cylindrical support region $\mathcal{N}_i$ centered around $\boldsymbol{x}_i$ are expressed in 2-D cylindrical coordinates (α, β), where α is the nonnegative perpendicular radial distance to $\boldsymbol{n}_i$ and β denotes the signed elevation component w.r.t. the surface tangent plane defined by $\boldsymbol{x}_i$ and $\boldsymbol{n}_i$, see Fig. 3c. The descriptor is then established as a 2-D histogram over the (α, β) space of $\mathcal{N}_i$. In the original formulation [13], the histogram bin width is derived from the median edge length of the surface mesh. However, this is not feasible for multi-modal surface registration entailing different mesh resolutions. Instead, we use a fixed metric bin width.

3.2. Correspondence Search Strategy

Our datasets (template data $\mathcal{T}$ and reference data $\mathcal{R}$) are represented as two sets of pairs of surface coordinates $\boldsymbol{x}$ and their associated feature descriptors $\boldsymbol{f}$:

$$\mathcal{T} = \left\{ \left(\boldsymbol{x}_i^{\mathcal{T}}, \boldsymbol{f}_i^{\mathcal{T}} \right) \right\}, \quad \mathcal{R} = \left\{ \left(\boldsymbol{x}_j^{\mathcal{R}}, \boldsymbol{f}_j^{\mathcal{R}} \right) \right\}, \qquad (1)$$

where $\boldsymbol{f} \in \mathbb{R}^D$ denotes a feature vector of dimensionality D. Given a template point $\boldsymbol{x}_i^{\mathcal{T}}$, the corresponding reference point $\boldsymbol{x}_{j^*}^{\mathcal{R}}$ is then determined by searching for the best match between the feature descriptors using a cross validation strategy with an appropriate similarity metric $\mathcal{S}$ (see Sec. 4.2):

$$j^* = \arg\min_j \mathcal{S}\left(\boldsymbol{f}_i^{\mathcal{T}}, \boldsymbol{f}_j^{\mathcal{R}} \right). \qquad (2)$$

For the purpose of eliminating false correspondences, the set of correspondences is pruned by applying a geometric consistency check [10]. Based on an iterative scheme, we successively penalize and remove matches that exhibit inconsistent surface normals and locations. The set of remaining correspondences $\mathcal{C} = \left\{ \left(\boldsymbol{x}_i^{\mathcal{T}}, \boldsymbol{x}_{j^*}^{\mathcal{R}} \right) \right\}$ is then used to estimate the rigid body transformation $(\boldsymbol{R}^*, \boldsymbol{t}^*)$:

$$(\boldsymbol{R}^*, \boldsymbol{t}^*) = \arg\min_{\boldsymbol{R}, \boldsymbol{t}} \frac{1}{|\mathcal{C}|} \sum_{\left(\boldsymbol{x}_i^{\mathcal{T}}, \boldsymbol{x}_{j^*}^{\mathcal{R}} \right) \in \mathcal{C}} \| (\boldsymbol{R}\boldsymbol{x}_i^{\mathcal{T}} + \boldsymbol{t}) - \boldsymbol{x}_{j^*}^{\mathcal{R}} \|_2^2$$

where $\boldsymbol{R} \in \mathbb{R}^{3\times3}$ denotes a rotation matrix and $\boldsymbol{t} \in \mathbb{R}^3$ a translation vector. In this work, $\boldsymbol{R}$ is restricted to the rotation about the table's vertical isocenter axis, since standard RT treatment tables are limited to four degrees of freedom (translation and rotation about one axis). The optimization problem is solved using a least-squares estimator.

4. Experiments

For quantitative evaluation of the proposed framework, we have benchmarked the performance of the descriptors on real data from a low-cost RI device (Microsoft Kinect). Indeed, the method is generic in a sense that it can be applied with various RI technologies. First, in an experimental study on anthropomorphic phantoms, we investigate the method's potential for multi-modal RI/CT registration. Thereby, we underline the benefits of the proposed method for partial matching. Second, we study the performance of the algorithm on data from healthy volunteers (RI/RI registration), where deformations may occur due to variations in patient body distortion and respiratory motion.

4.1. Methods and Materials

Benchmark Dataset. We have generated a database[1] of Kinect RI data (640×480 pixels) for two anthropomorphic phantoms (male/female) and three healthy volunteers. Data were acquired in a clinical radiation therapy environment (Siemens ARTISTE). For each phantom (volunteer), we have captured RI data for $N = 20$ (4) different initial misalignments of the treatment table, including large deviations of up to 200 mm and 45°. The set of poses for the phantom benchmark is composed of all possible permutations of the transformation parameter sets $\varphi = \{0, 5, 10, 25, 45\}^\circ$, $t_{\text{SI}} = \{0, 200\}$ mm, and $t_{\text{ML}} = \{0, 200\}$ mm, where the angle φ describes the table rotation about the isocenter axis and $t_{\text{SI}}, t_{\text{ML}}$ denote the table translation in superior-inferior (SI) and medio-lateral (ML) direction. The translation in anterior-posterior (AP) direction was set to $t_{\text{AP}} = -600$ mm, representing the initial height for the patient to get on the table and recline. For the volunteer study, the space of transformations was $\varphi = \{0, 10\}^\circ$ and $t_{\text{SI}} = \{0, 200\}$. The table positioning control (accuracy: ± 1.0 mm, ± 0.5°) was used to set up the respective ground truth transformation $(\boldsymbol{R}_{\text{GT}}, \boldsymbol{t}_{\text{GT}})$. The RI sensor was mounted 200 cm above the floor, at a distance of 240 cm to the LINAC isocenter and a viewing angle of 55°. CT data of the phantoms were acquired on a Siemens SOMATOM scanner.

Data Preprocessing. The patient surface is extracted from CT data using a thresholding based region growing segmentation and a marching cubes algorithm on the resulting binary segmentation mask followed by Laplacian mesh smoothing. Subsequently, the mesh was decimated in order to reduce the computational complexity. Let us remark

[1] CT/RI data and corresponding ground truth table transformations are available from the authors for noncommercial research purposes, serving as a baseline for benchmarks with future competing approaches.

Descriptor	**SR**	Error	(m)	(f)	**(m) & (f)**	$\varphi_{\mathrm{GT}} \leq 10^\circ$	$25^\circ \leq \varphi_{\mathrm{GT}}$
MeshHOG	**0.98**	$\Delta\varphi[^\circ]$	1.0 ± 0.6	2.0 ± 1.6	$\mathbf{1.5 \pm 1.3}$	1.4 ± 1.0	1.7 ± 1.9
		$\Delta\boldsymbol{t}[mm]$	13.7 ± 7.0	13.1 ± 5.4	$\mathbf{13.4 \pm 6.2}$	12.6 ± 3.9	14.6 ± 6.9
RIFF	**0.95**	$\Delta\varphi[^\circ]$	1.4 ± 1.2	2.0 ± 1.6	$\mathbf{1.7 \pm 1.4}$	1.3 ± 1.2	2.3 ± 1.8
		$\Delta\boldsymbol{t}[mm]$	11.0 ± 4.1	12.8 ± 6.0	$\mathbf{11.8 \pm 5.1}$	12.2 ± 5.7	11.3 ± 3.7
Spin images	**0.95**	$\Delta\varphi[^\circ]$	0.7 ± 0.6	0.6 ± 0.5	$\mathbf{0.7 \pm 0.6}$	0.7 ± 0.5	0.7 ± 0.4
		$\Delta\boldsymbol{t}[mm]$	13.3 ± 6.0	12.1 ± 4.8	$\mathbf{12.7 \pm 5.4}$	11.8 ± 3.2	14.0 ± 5.8

Table 1. Mean rotational and translational errors for multi-modal RI/CT surface registration on male (m) and female (f) anthropometric phantoms. SR quotes the percentage of successful registrations, classified with heuristic thresholds ($\Delta\varphi < 10^\circ$, $\Delta\boldsymbol{t} < 40$ mm). The two columns on the right oppose the combined results of the male and female phantom on small and large rotations.

that CT preprocessing can be performed offline prior to the first fraction. In order to improve the SNR of the RI measurements, we combine temporal averaging (over 150 ms) with edge-preserving filtering, invalid range measurements are restored using normalized convolution [24]. The patient surface can be segmented from the background by incorporating prior information about the treatment table plane. The preprocessed RI (CT) meshes consist of $\sim$15k (20k) vertices. Note that CT data typically covers only a portion of the RI scene.

4.2. Results and Discussion

In order to assess the accuracy of the method, we have registered the RI dataset of $N = 20$ phantom poses (see Sec. 4.1) to the phantom's CT surface. The CT data was previously aligned to an RI reference at isocenter position ($\varphi = 0^\circ, t_{\mathrm{SI}} = 0$ mm, $t_{\mathrm{ML}} = 0$ mm, $t_{\mathrm{AP}} = -150$ mm) using landmarks. We then compare the estimated table transformation ($\boldsymbol{R}^*, \boldsymbol{t}^*$) to the ground truth table setup ($\boldsymbol{R}_{\mathrm{GT}}, \boldsymbol{t}_{\mathrm{GT}}$) by computing the mean rotational and mean translational errors over the set of N poses:

$$\Delta\varphi = \frac{1}{N}\sum_{i=1}^{N} |\varphi_i^* - \varphi_{i,\mathrm{GT}}|, \quad \Delta\boldsymbol{t} = \frac{1}{N}\sum_{i=1}^{N} \|\boldsymbol{t}_i^* - \boldsymbol{t}_{i,\mathrm{GT}}\|_2$$

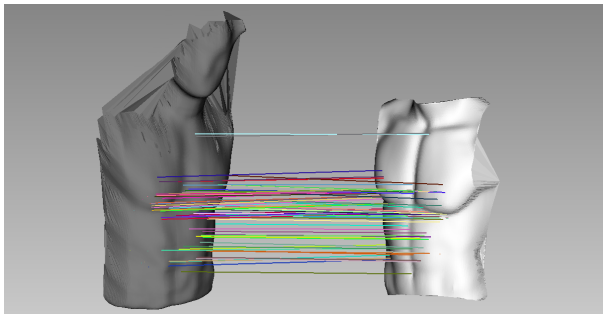
where φ^* (φ_{GT}) denotes the estimated and ground truth rotation angle about the table axis, $\boldsymbol{t}^*$ ($\boldsymbol{t}_{\mathrm{GT}}$) the translation. For this case study, the descriptor parameters were set to typical values [13, 22, 25]. To achieve good repeatability despite of noise and quantization effects in the RI data, the support region radius was set to $r_{\mathcal{N}} = 100$ mm. For the similarity metric $\mathcal{S}$, we used a correlation distance metric for spin images and the L^1-norm for MeshHOG and RIFF, respectively, neglecting classical measures for comparison of density distributions (e.g. Kullback-Leibler divergence) to limit the computational effort. Prior to matching, the MeshHOG and RIFF descriptors were L^2-normalized in order to cope with different mesh resolutions. A qualitative illustration of the extracted set of correspondences is shown in Fig. 4. Quantitative results for multi-modal RI/CT registration are depicted in Table 4.1. For all three descriptors, the registration framework was able to estimate the table transformation for the vast majority of initial misalignments. For the application in RT patient setup, what is most important is a high percentage of successful registrations (SR). Extreme accuracy is not essential for this initial alignment. Having achieved the highest percentage of successful registrations (97.5%), let us refer to the results of the MeshHOG descriptor as an overall performance indicator yielding a mean rotational and translational error of $\Delta\varphi = 1.5 \pm 1.3^\circ$ and $\Delta\boldsymbol{t} = 13.4 \pm 6.2$ mm (RI/CT), respectively. The achieved level of accuracy is consistent with manual setup using lasers and skin markers that is clinical practice today. Please note that spin images slightly outperformed the other descriptors in terms of accuracy ($\Delta\varphi = 0.7 \pm 0.6^\circ$, $\Delta\boldsymbol{t} = 12.7 \pm 5.4$ mm), but the low SR rate makes them less appropriate for this task. Let us remark that the scope of this paper is restricted to a coarse initial patient setup. Setup verification in terms of accurate positioning refinement is a mandatory second step but not addressed here.

Resilience to Deformations In a volunteer study, we have investigated the resilience of the proposed method w.r.t. variations in surface topology. Here, we have captured RI data from three volunteers at arbitrary states within the respiration cycle for the reduced set of 4 benchmark poses. These data are registered (mono-modal) to an RI reference at isocenter position ($\varphi = 0^\circ$, $t_{\mathrm{SI}} = 0$ mm, $t_{\mathrm{ML}} = 0$ mm, $t_{\mathrm{AP}} = -150$ mm) and quantitatively evaluated. All four poses were successfully registered. The results of $\Delta\varphi = 1.3 \pm 0.9^\circ$ and $\Delta\boldsymbol{t} = 15.0 \pm 1.0$ mm (MeshHOG) indicate that the modified surface descriptors and our correspondence search engine are capable of coping with minor deformations due to body distortion and respi-

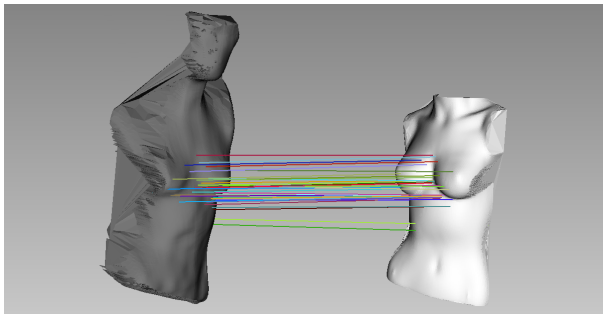
ratory motion. Let us note that respiratory motion typically evokes a thoracic AP movement in the scale of $\sim$10 mm for regular breathing [21].

5. Conclusions

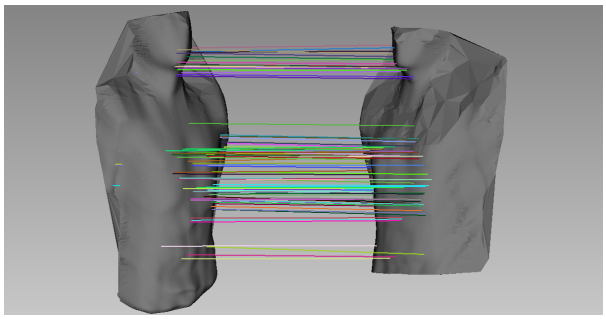
We have presented a novel markerless solution for the automation of the initial coarse patient setup in RT. Based on a multi-modal registration of RI and CT data, the approach renders the conventional initialization using lasers and skin markers redundant. On real data from Microsoft Kinect, we achieved an accuracy of $\Delta\varphi = \pm 1.5^\circ$ and $\Delta \boldsymbol{t} = \pm 13.4$ mm at an SR rate of 97.5% for anthropometric phantoms, providing a reliable initialization for subsequent refinement with verification systems. Our modified MeshHOG descriptor makes the method more robust, outperforming the two other proposed descriptors regarding the percentage of successful registrations. Experiments on volunteer data have substantiated the framework's capability of coping with deformations that occur due to body distortion and respiratory motion. Further investigations concerning multi-scale descriptor representations and a setup with multiple RI cameras, providing an increased coverage of the patient surface, will be subject of our upcoming research.



(a)



(b)



(c)

Figure 4. (a), (b): Spatial distribution of point correspondences for a multi-modal RI/CT registration (male and female phantoms). Note the partial matching issue and the concentration of correspondences in regions with salient surface topology. (c): Point correspondences for a mono-modal RI/RI registration (volunteer data). For convenience, only a subset of the found correspondences is shown. Triangulation issues in the upper thoracic region result from the flat viewing angle of the RI sensor.

Acknowledgments

S. Bauer and J. Wasza gratefully acknowledge the support by the European Regional Development Fund (ERDF) and the Bayerisches Staatsministerium für Wirtschaft, Infrastruktur, Verkehr und Technologie (StMWIVT), in the context of the R&D program IuK Bayern under Grant No. IUK338. S. Haase is supported by the Deutsche Forschungsgemeinschaft (DFG) under Grant No. HO 1791/7-1. The authors further acknowledge the support of Prof. Dörfler (Department of Neuroradiology, Erlangen University Clinic, Germany) in acquiring phantom CT data.

References

[1] S. Bauer, B. Berkels, J. Hornegger, and M. Rumpf. Joint ToF image denoising and registration with a CT surface in radiation therapy. In *Proceedings of International Conference on Scale Space and Variational Methods in Computer Vision*, volume 6667 of *LNCS*, pages 98–109. Springer, May 2011.

[2] C. Bert, K. G. Metheany, K. Doppke, and G. T. Y. Chen. A phantom evaluation of a stereo-vision surface imaging system for radiotherapy patient setup. *Medical Physics*, 32(9):2753–2762, Sep 2005.

[3] J. Besl and M. Neil. A method for registration of 3-D shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(2):239–256, 1992.

[4] A. Brahme, P. Nyman, and B. Skatt. 4D laser camera for accurate patient positioning, collision avoidance, image fusion and adaptive approaches during diagnostic and therapeutic procedures. *Medical Physics*, 35(5):1670–1681, 2008.

[5] B. Bustos, D. A. Keim, D. Saupe, T. Schreck, and D. V. Vranić. Feature-based similarity search in 3D object databases. *ACM Computing Surveys*, 37:345–387, 2005.

[6] C. S. Chua and R. Jarvis. Point signatures: A new representation for 3D object recognition. *International Journal of Computer Vision*, 25:63–85, 1997.

[7] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, volume 1, pages 886–893, 2005.

[8] T. Frenzel. Patient setup using a 3D laser surface scanning system. In *Proceedings of IFMBE World Congress on Medical Physics and Biomedical Engineering*, volume 25/1, pages 217–220. Springer, Sep 2009.

[9] A. Frome, D. Huber, R. Kolluri, T. Bülow, and J. Malik. Recognizing objects in range data using regional point descriptors. In *Proceedings of European Conference on Computer Vision*, pages 224–237. Springer, May 2004.

[10] T. Funkhouser and P. Shilane. Partial matching of 3D shapes with priority-driven search. In *Proceedings of Eurographics Symposium on Geometry Processing*, pages 131–142, 2006.

[11] R. Gal and D. Cohen-Or. Salient geometric features for partial shape matching and similarity. *ACM Transactions on Graphics*, 25:130–150, Jan 2006.

[12] D. P. Gierga, M. Riboldi, J. C. Turcotte, G. C. Sharp, S. B. Jiang, A. G. Taghian, and G. T. Chen. Comparison of target registration errors for multiple image-guided techniques in accelerated partial breast irradiation. *International Journal of Radiation Oncology Biology Physics*, 70(4):1239–1246, 2008.

[13] A. E. Johnson and M. Hebert. Using spin images for efficient object recognition in cluttered 3D scenes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21:433–449, May 1999.

[14] M. Krengli, S. Gaiano, E. Mones, A. Ballar, D. Beld, C. Bolchini, and G. Loi. Reproducibility of patient setup by surface image registration system in conformal radiotherapy of prostate cancer. *Radiation Oncology*, 4:9, 2009.

[15] P. Kupelian et al. Multi-institutional clinical experience with the calypso system in localization and continuous, real-time monitoring of the prostate gland during external radiotherapy. *International Journal of Radiation Oncology Biology Physics*, 67(4):1088–1098, 2007.

[16] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60:91–110, Nov 2004.

[17] K. Müller, S. Bauer, J. Wasza, and J. Hornegger. Automatic multi-modal ToF/CT organ surface registration. In *Proceedings of Bildverarbeitung für die Medizin*, pages 154–158. Springer, Mar 2011.

[18] J. L. Peng, D. Kahler, J. G. Li, S. Samant, G. Yan, R. Amdur, and C. Liu. Characterization of a real-time surface image-guided stereotactic positioning system. *Medical Physics*, 37(10):5421–5433, Oct 2010.

[19] S. Rusinkiewicz and M. Levoy. Efficient variants of the ICP algorithm. In *Proceedings of International Conference on 3-D Digital Imaging and Modeling*, pages 145–152, 2001.

[20] P. J. Schöffel, W. Harms, G. Sroka-Perez, W. Schlegel, and C. P. Karger. Accuracy of a commercial optical 3D surface imaging system for realignment of patients for radiotherapy of the thorax. *Physics in Medicine and Biology*, 52(13):3949–3963, Jul 2007.

[21] W. Segars, S. Mori, G. Chen, and B. Tsui. Modeling respiratory motion variations in the 4D NCAT phantom. In *Proceedings of IEEE NSS/MIC*, volume 4, pages 2677–2679, 2007.

[22] G. Takacs, V. Chandrasekhar, S. Tsai, D. Chen, R. Grzeszczuk, and B. Girod. Unified real-time tracking and recognition with rotation-invariant fast features. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 934–941, 2010.

[23] F. Tombari, S. Salti, and L. Di Stefano. Unique signatures of histograms for local surface description. In *Proceedings of European Conference on Computer Vision*, pages 356–369. Springer, 2010.

[24] 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 *Proceedings of International Conference on Computer Vision, IEEE Workshop on Consumer Depth Cameras for Computer Vision*, Nov 2011.

[25] A. Zaharescu, E. Boyer, K. Varanasi, and R. P. Horaud. Surface feature detection and description with applications to mesh matching. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 373–380, 2009.