

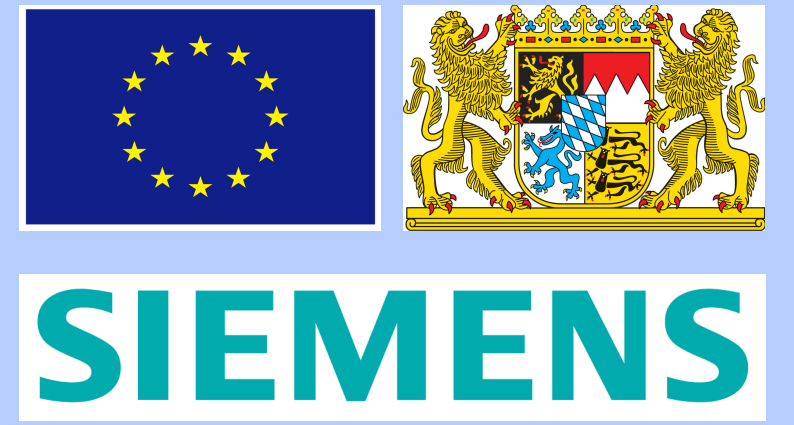
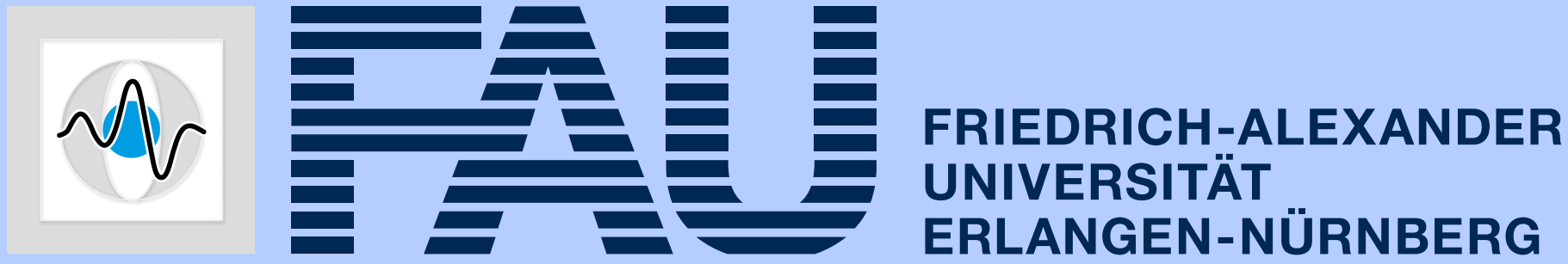
# Multi-modal Surface Registration for Markerless Initial Patient Setup in Radiation Therapy using Microsoft's Kinect Sensor

S. Bauer<sup>1</sup>, J. Wasza<sup>1</sup>, S. Haase<sup>1</sup>, N. Marosi<sup>3</sup>, J. Hornegger<sup>1,2</sup>

<sup>1</sup> Pattern Recognition Lab, Dept. of Computer Science

<sup>2</sup> Erlangen Graduate School in Advanced Optical Technologies (SAOT)  
Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany

<sup>3</sup> Siemens AG, Healthcare Sector, Erlangen, Germany



## Coarse Initial Patient Setup in Fractionated Radiation Therapy

- **Patient setup reproducibility in fractionated radiation therapy (RT):** Prior to each fraction, the oncological patient must be aligned to reference planning data
- **Conventional coarse initial setup:** Manual procedure using lasers + skin markers → tedious, time-consuming and error-prone
- **Non-radiographic setup verification using range imaging (RI) modalities:**
  - Automatic patient (re-)positioning at a fine scale
  - Requires conventional coarse initial alignment

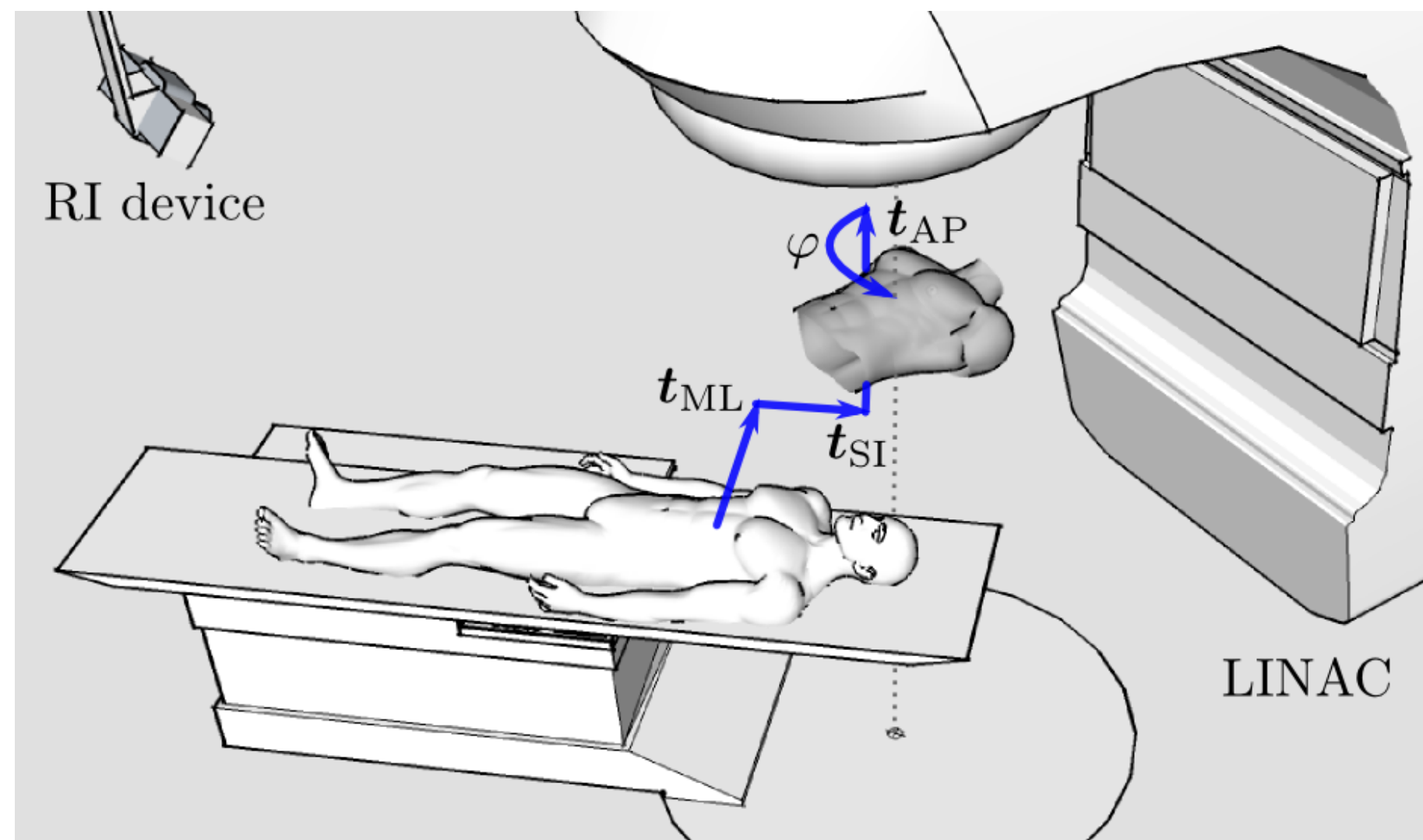


Fig. 1: Illustration of the proposed automatic coarse initial patient setup. The estimated transformation (blue) is applied to the treatment table.

- **Goal: Markerless RI-based solution** for automatic coarse initial patient setup → superseding the need for lasers and skin markers
- **Idea:** Estimate transformation from correspondences based on matching local surface descriptors → capable of coping with partial matching → capable of handling gross initial misalignments
- **Challenge:** Multi-modal surface description (CT/RI) → invariance to mesh representation and resolution → robustness to variations in topology that occur due to body distortion and respiratory motion

## Surface Registration Framework

- **Point-based registration method matching local invariant descriptors:**

1. Encode local topology by invariant features → sets of coordinate/feature pairs
2. Establish correspondences by descriptor matching (cross validation strategy)
3. Prune set of correspondences by geometric consistency analysis
4. Estimate rigid-body transformation w.r.t. treatment table coordinate system

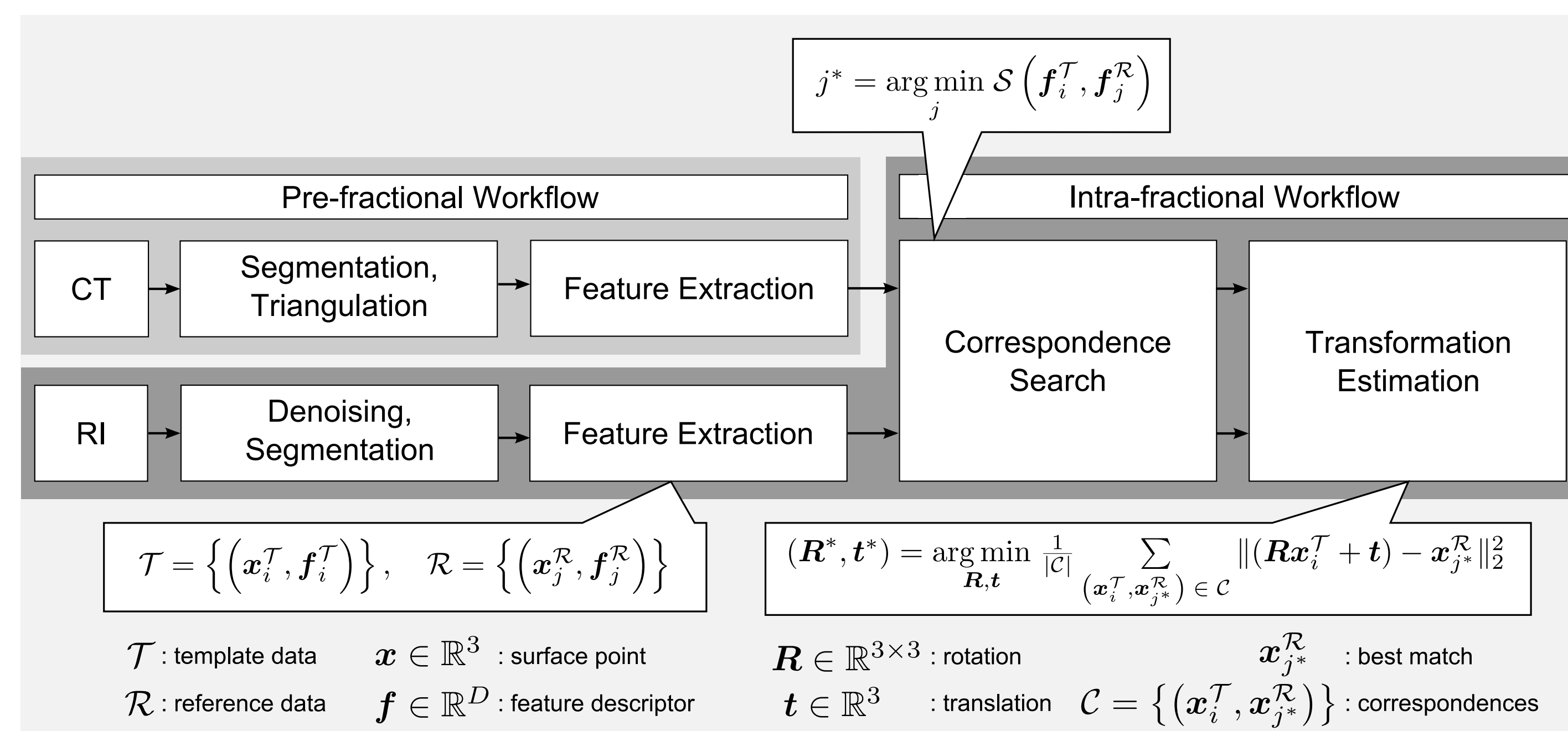
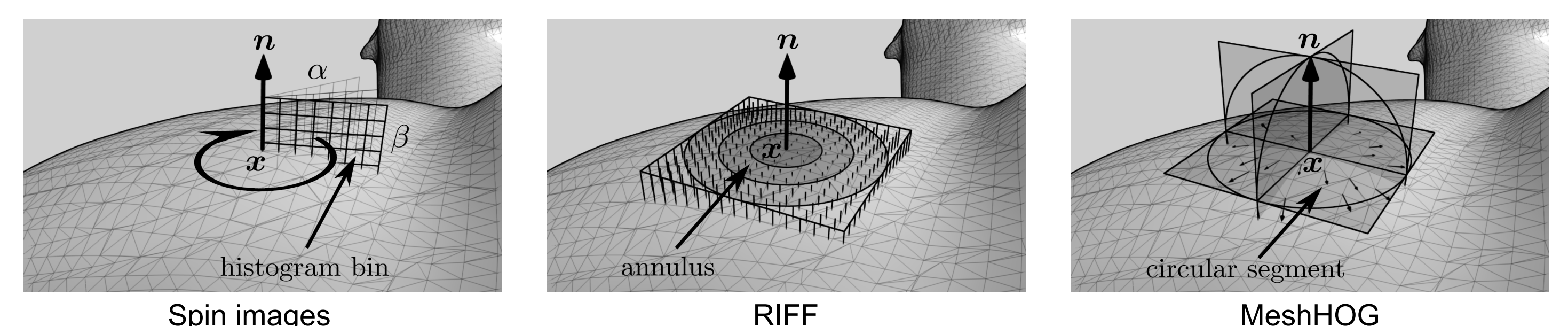


Fig. 2: Flowchart of the feature-based surface registration framework. Data propagates from left to right. The shading indicates the parts of the workflow that are performed prior to the first fraction.

## 3-D Surface Descriptors

- **Contribution:** Extension of state-of-the-art descriptors to handle different mesh resolutions (multi-modal registration scenario) and topological variations that occur due to low signal-to-noise ratio of RI sensors.

Histograms trade-off descriptive power and positional precision for robustness and repeatability by compressing geometry into bins:



- **Baseline: Spin images** [1]: Fixed metric bin width for multi-modal application
- **RIFF** [2]: Orthographic depth representation of the local 3-D surface topology w.r.t. tangent plane in a planar patch → 2-D descriptors can be applied to surface data
- **MeshHOG** [3]: Extension of HOG for scalar fields defined on surface manifolds → Circular uniform surface sampling [4] to cope with arbitrary mesh resolutions → Scalar field: signed distance to local regression plane instead of curvature

## Experiments and Results

- **Case study on real data from low-cost RI device (Microsoft Kinect)**
  - Data acquired in a clinical RT environment (Siemens ARTISTE)
  - RI sensor: 240 mm from LINAC isocenter, 55° viewing angle, preprocessing: defect interpolation, temporal averaging (150 ms), edge-preserving filtering
- **Evaluation:** compare estimated transformation to ground truth for set of N poses:
 
$$\Delta\varphi = \frac{1}{N} \sum_{i=1}^N |\varphi_i^* - \varphi_{i,GT}|, \quad \Delta t = \frac{1}{N} \sum_{i=1}^N \|t_i^* - t_{i,GT}\|_2 \quad \varphi \in \mathbb{R} : \text{rotation angle} \quad t \in \mathbb{R}^3 : \text{translation}$$
- **Multi-modal registration:**
  - Male (m) and female (f) anthropomorphic phantoms (CT/RI data)
  - N = 20 different large initial misalignments [0, 200] mm, [0, 45]°
- **Mono-modal registration:**
  - RI data from three healthy volunteers, captured at arbitrary respiration states
  - N = 4 different large initial misalignments [0, 200] mm, [0, 10]°

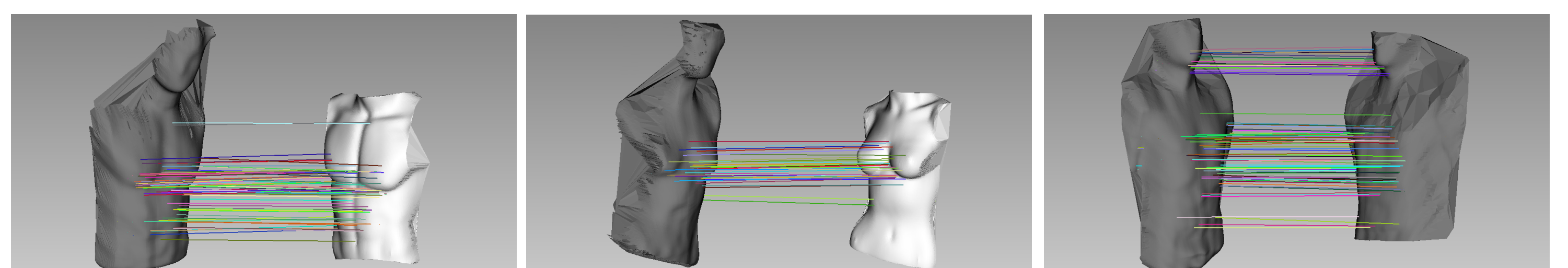


Fig. 3: Spatial distribution of point correspondences for multi-modal CT/RI (left, middle), for male and female phantom, and mono-modal registration (RI/RI), for volunteer data. Only a subset of the found correspondences is shown. Note the partial matching issue and the concentration of correspondences in regions with salient surface topology.

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

## Conclusions

- **Modified surface descriptors capable of coping with different mesh resolutions and variations in topology**
- Feasible coarse initialization for patient setup refinement with RI-based position verification systems in RT
- Outlook: Multi-scale descriptor representations, multi-camera setup

**Acknowledgments:** S. Bauer and J. Wasza acknowledge the support by the European Regional Development Fund (ERDF) and the Bayerisches Staatsministerium für Wirtschaft, Infrastruktur, Verkehr und Technologie (StMWIVT) under grant no. IUK338/001. S. Haase acknowledges support by the Deutsche Forschungsgemeinschaft (DFG) under grant no. HO 1791/7-1. The authors further acknowledge the support of Prof. Dörfler (Erlangen University Clinic, Germany) in acquiring phantom CT data.

## References

- [1] A. E. Johnson and M. Hebert. Using spin images for efficient object recognition in cluttered 3D scenes. IEEE TPAMI, 21:433-449, 1999.
- [2] 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 Proc IEEE CVPR, p. 934-941, 2010.
- [3] A. Zaharescu, E. Boyer, K. Varanasi, and R. P. Horaud. Surface feature detection and description with applications to mesh matching. In Proc IEEE CVPR, p. 373-380, 2009.
- [4] S. Bauer, J. Wasza, K. Müller, J. Hornegger. 4-D Photogeometric face recognition with Time-of-Flight sensors. In Proc IEEE WACV, p. 196-203, 2011.