International Journal of Computer Assisted Radiology and Surgery manuscript No.
(will be inserted by the editor)

Hybrid electromagnetic and image-based tracking of endoscopes with guaranteed smooth output

Tobias Reichl · Xiongbiao Luo · Manuela Menzel · Hubert Hautmann · Kensaku Mori · Nassir Navab

Received: date / Accepted: date

Abstract Purpose: Flexible fiberoptic bronchoscopy is a widespread medical procedure for diagnosis and treatment of lung diseases. Navigation systems are needed to track the flexible endoscope within the bronchial tree. Electromagnetic (EM) tracking is currently the only technology used clinically for this purpose. The registration between EM tracking and patient anatomy may become inaccurate due to breathing motion, so the addition of image-based tracking has been proposed as a hybrid EM-image based system.

Materials & Methods: When EM tracking is used as an initialization for image registration, small changes in the initialization may lead to different local minima, and noise is amplified by hybrid tracking. The tracking output is modeled as continuous and uses splines for interpolation, thus smoothness is greatly improved. The bronchoscope pose relative to Computed Tomography (CT) data is interpolated using Catmull-Rom splines for position and spherical linear interpolation (SLERP) for orientation.

Results: The hybrid method was evaluated using ground truth poses manually selected by experts, where mean inter-expert agreement was determined as 1.26 mm. Using four dynamic phantom data sets, the accuracy was 4.91 mm, which is equivalent to previous methods. Compared to state-of-art methods, inter-frame smoothness was improved from 2.77-3.72 mm to 1.24 mm.

Conclusion: Hybrid image and electromagnetic endoscope guidance provides a more realistic and physically plausible solution with significantly less jitter. This quantitative result is confirmed by visual comparison of real and virtual video, where the virtual video output is much more consistent and ro-

T. Reichl and N. Navab
Computer-Aided Medical Procedures, TUM, München, Germany E-mail: reichl@tum.de

X. Luo and K. Mori
Information and Communications Headquarters, Nagoya University, Japan

M. Menzel and H. Hautmann
Medizinische Klinik I, Klinikum rechts der Isar, TUM, München, Germany

bust, with fewer occasions of tracking loss or unexpected movement compared with previous methods.

1 Introduction

Lung cancer is the leading cause of cancer death worldwide [22]. Flexible bronchoscopy is a routine procedure for diagnosis and treatment of lung diseases, and one of its most common applications is transbronchial biopsy (TBB) of suspected lung lesions. TBB is commonly performed after lung lesions have been identified on CT images, and it is definitely beneficial to transfer this three-dimensional (3-D) information to the operating room. Bronchoscopy is a monitor-based procedure, and augmentation of video images with overlay images for guidance or targeting is straightforward. This promises high clinical acceptance due to smooth integration into the clinical workflow.

Due to line of sight issues, it is not possible to use optical tracking. Also, due to the flexibility of the instrument, it is not possible to infer precise information about the position and orientation of the endoscope tip from parts outside the body. This leaves electromagnetic (EM) tracking as the method of choice, which has also been clinically established for navigated bronchoscopy [44, 17, 40], and which is used in the commercially available iLogic bronchoscopy navigation system (superDimension, Minneapolis, MN, USA).

However, EM tracking only measures position and orientation in space, i.e. not directly in relation to patient anatomy. The registration between EM tracking and patient anatomy may become inaccurate due to breathing motion, therefore the combination with image-based tracking was proposed by Mori *et al.* [31]. Image-based tracking is less affected by breathing motion, since the endoscope is moving together with the surrounding anatomy. For image-based tracking, a cost function is minimized, which corresponds to the differences between the real endoscope image and the CT image, and the result of the optimization is usually a local minimum of this cost function.

Hybrid tracking complements the advantages of image-based tracking and EM tracking, since image-based tracking is inaccurate for frames with little to no information due to low contrast. EM tracking suffers from noisy artifacts while the sensor is moved [21], and relative errors due to patient respiration or heartbeat. However, with a hybrid method, not only noise, but also imperfect translation is propagated. From these limitations, the hybrid tracking method was further improved, e.g. by Deguchi *et al.* [9], Soper *et al.* [45], and Luo *et al.* [27, 28].

Related Tracking Technologies: Besides EM tracking there are a number of alternative tracking technologies, as discussed below.

Inertial Tracking: Certain kinds of gyroscopes may be used for inertial tracking. Inertial sensors may be very small, like electromagnetic tracking sensors, and can be unobtrusively integrated in instruments, e.g. for endoscopic image re-orientation [19], but due to the integration of measurements, even small errors lead to a strong drift, which needs intermittent correction.

Transponder Localisation: New developments in radio frequency identification (RFID) technology enable millimetre accuracy localisation of RFID tags using phase difference [18, 51]. Due to its wireless nature, the inherent identification capability, and the extremely small size of RFID tags, this technology is very promising, but has not yet been applied in clinical environments.

The "GPS for the Body" is a wireless localisation system (Calypso, Seattle, WA, USA). With this system, a set of three transponders (8.5 mm length) is used to track a single target's position. For real-time organ tracking in prostate radiotherapy an error of 2.0 mm or better was determined [4, 52, 25], and integration into an augmented reality system was proposed [34]. In another system an implanted permanent magnet can be localised by sensors outside the body [41].

Fibre-Optic Tracking: Owing to modern optics, it is possible to measure strain and temperature quasi-continuously along an optical fibre. Commercially available reflectometers (Luna Technologies, Blacksburg, VA, USA) can measure 7 million data points along a 70 m long fibre, which corresponds to a spatial resolution of 10 μm. The commercially available "ShapeTape" device (Measurand, Fredericton, Canada) has already been used to track a two-dimensional (2-D) ultrasound probe for freehand 3-D ultrasound [36, 20], for endoscopic 3-D ultrasound systems [24], or for measurement of body-seat interface shape [26], but accuracy was too low to be useful [24, 3].

Smooth Hybrid Electromagnetic and Image-Based Tracking: In general, any rigid registration between CT and EM tracking coordinate systems will become inaccurate due to breathing motion. In order to track the endoscope motion relative to patient anatomy, these dynamic inaccuracies need to be corrected and image-based tracking methods have been presented. While the common objective has been optimal accuracy, with hybrid EM and image-based tracking methods small changes in the EM measurements can also lead to large jumps between local minima after image-based registration. Such jumps in the tracked endoscope position and orientation affect where augmented reality overlays are placed, leading to a shaky visualization. It has been recognized in the augmented reality scientific community, e.g. by Azuma and Bishop [2], that such high-frequency jitter appear to users as objectionable oscillations, which do not correspond to the user's actual motion and may lower performance and user acceptance. Smoothing the input signal is ineffective, since low-pass filtering adds latency. A new approach to this problem is modeling the true trajectory as continuous. This will yield a more consistent, physically plausible, and clinically acceptable tracking result. While a preliminary description of this method has been presented before [39], here we present it in greater detail with more context and discussion.

In the following, we are going to present how we create virtual images from the patient CT data, how we use these for hybrid electromagnetic and image-based tracking, and how we enforce smoothness and continuity of the output.

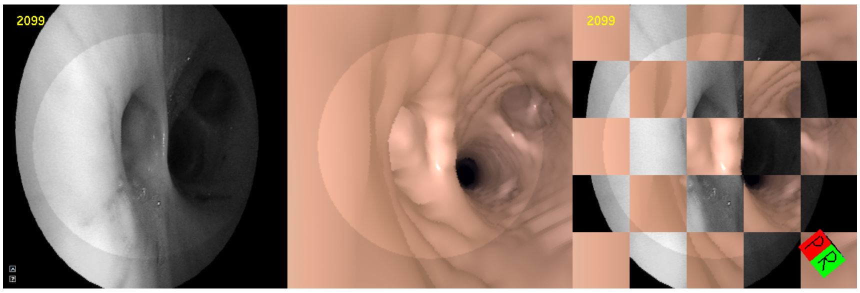

Fig. 1 Navigated bronchoscopy user interface with real bronchoscopic image (left), virtual bronchoscopic image rendered at current position and orientation (centre), and chequerboard overlay (right).

While the main contribution of our method is an improvement of smoothness, we also evaluate the accuracy of our method with respect to ground truth data from human experts, in order to demonstrate that the improved smoothness is not achieved at the cost of reduced accuracy. Then, we discuss our method and results and we relate the proposed method to the state of the art in the scientific community.

2 Methods

Image-based endoscope tracking can be considered as intensity-based registration of multiple 2-D video images and one 3-D CT volume. There has been a large body of work on 2-D/3-D registration, mostly considering registration of 3-D CT or Magnetic Resonance Imaging (MRI) data to intra-operative 2-D X-ray imaging [29]. For bronchoscopy, so-called *virtual bronchoscopy* has been proposed, and may be used for 2-D/3-D registration.

For virtual bronchoscopy visualization of any CT data sets we use an iso-surface renderer to compute a bronchoscopy-like image on the graphics processing unit (GPU), as shown in Fig. 1. For each pixel, from the position of the virtual camera a number of rays is cast in the viewing direction, and the nearest intersection with a given threshold value is determined and rendered. Due to the interpolation capabilities of the GPU, the resulting surface is smooth and closely corresponds to a surface obtained by an airway segmentation by region growing [32,43,48].

As already introduced in the previous work [39], for smooth hybrid EM and image-based tracking we need to model the real bronchoscope motion as continuous. When electromagnetic tracking is used as an initialization for image registration, small changes in the initialization may lead to different local minima, and noise is amplified by hybrid tracking. By modeling the output as continuous and using splines for interpolation, smoothness is greatly improved.

Such a continuous description of bronchoscope pose at time t is given by its position $p(t)$ and orientation $q(t)$. Since both the real movement of the bronchoscope and the movement of anatomy (breathing, heartbeat, etc.) are spatially smooth over time, movement of the bronchoscope relative to patient anatomy is smooth as well.

Control points are assigned to frames with a pre-determined regular time spacing s, e.g. one control point every five frames, and initialised with the original EM tracking measurements. For interpolation of position we use Catmull-Rom splines [7], a special class of cubic Hermite splines, where tangents are continuous over multiple segments. The resulting curve is continuously differentiable and passes directly through the control points:

$$\mathbf{p}(t) = \frac{1}{2} \begin{pmatrix} 1 & u & u^2 & u^3 \end{pmatrix} \begin{pmatrix} 0 & 2 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ 2 & -5 & 4 & -1 \\ -1 & 3 & -3 & 1 \end{pmatrix} \begin{pmatrix} \mathbf{p}_{i-1} \\ \mathbf{p}_i \\ \mathbf{p}_{i+1} \\ \mathbf{p}_{i+2} \end{pmatrix}, \tag{1}$$

where $\mathbf{p}_{i-1\ldots i+2}$ are positions of consecutive control points, $\lfloor \cdot \rfloor$ is the floor operator, $i = \lfloor t/s \rfloor$ is a control point index, and $u = t/s - \lfloor t/s \rfloor$ is the interpolation ratio between control points $\mathbf{p}_i$ and $\mathbf{p}_{i+1}$.

For orientation we use quaternions, because they allow a continuous representation without gimbal lock. Then, for interpolation between quaternions we use spherical linear interpolation (SLERP) [42], which provides "smooth and natural" motion:

$$\mathbf{q}(t) = \mathbf{q}_i \, (\mathbf{q}_i^{-1} \, \mathbf{q}_{i+1})^u = \frac{\sin(1-u)\theta}{\sin\theta} \, \mathbf{q}_i + \frac{\sin u\theta}{\sin\theta} \, \mathbf{q}_{i+1}, \tag{2}$$

where θ is the rotation difference between $\mathbf{q}_i$ and $\mathbf{q}_{i+1}$. Initial parameters $\mathbf{p}_k$ and $\mathbf{q}_k$ for all control points are taken directly from the EM tracking measurements.

Then, for each control point i its position $\mathbf{p}_i$ and orientation $\mathbf{q}_i$ are optimized with respect to a cost function $E(\mathbf{p}, \mathbf{q})$. For the matching the 2-D bronchoscope images and the 3-D CT data set, we perform volume rendering of the CT data and create a virtual bronchoscopic image $I_V(\mathbf{p}_k, \mathbf{q}_k)$ in terms of a camera with position $\mathbf{p}_k$ and orientation $\mathbf{q}_k$.

Adding regularizing measures, which incorporate EM tracking data and bending forces along the trajectory, we seek to maximize

$$E(\mathbf{p}, \mathbf{q}) = \underbrace{S(\mathbf{p}, \mathbf{q})}_{\text{similarity}} - \underbrace{\lambda_1 \cdot R_T(\mathbf{p}, \mathbf{q})}_{\text{displacement}} - \underbrace{\lambda_2 \cdot R_B(\mathbf{p})}_{\text{bending}}, \tag{3}$$

where λ_1 and λ_2 are weighting parameters. The cost function $E(\mathbf{p}, \mathbf{q})$ is only evaluated for control points, but for each control point, data from all the frames in its support is aggregated and used for updating the control point. Frames outside its support are not influenced by movement of this control point, so computation of the cost function can be decoupled for each control point. In addition, since the total number of frames is constant, the total computational effort does not depend on the number of control points.

The similarity $S(\mathbf{p},\mathbf{q})$ between real and virtual images is computed via the local normalized cross correlation (LNCC)

$$S(\mathbf{p},\mathbf{q}) = \sum_k \mathrm{LNCC}(I_R(t_k), I_V(\mathbf{p}_k,\mathbf{q}_k)), \tag{4}$$

where k is an image index in the neighborhood of the current control point, $I_R(t_k)$ is the real image at time t_k, and $I_V(\mathbf{p}_k,\mathbf{q}_k)$ is the virtual bronchoscopic image. The patch size for LNCC was set to 11×11 pixels. The gradient of $S(\mathbf{p},\mathbf{q})$ for each control point i can be approximated from the frames in its support $(i-2, i+2)$ via finite differences. This is closely related to bundle adjustment, where multiple images are registered for fiber-optic video mosaicking [1].

In order to penalize image artifacts, we penalize displacement from the EM tracking measurements trough a spring model. According to Hooke's law, spring force is proportional to displacement, so tension is

$$R_T(\mathbf{p},\mathbf{q}) = \sum_k \left[\alpha \cdot \|\mathbf{p}_k - \mathbf{p}_{k,0}\| + \theta(\mathbf{q}_k, \mathbf{q}_{k,0})\right], \tag{5}$$

where $\mathbf{p}_{k,0}$ and $\mathbf{q}_{k,0}$ are the position and orientation measured by EM tracking at time t_k, and the rotation difference $\theta(\mathbf{q}_k, \mathbf{q}_{k,0})$ is computed as the angular component of the difference quaternion. α is the ratio between rotational and translational spring constants and was set to $3.10°/\mathrm{mm}$, i.e. the ratio of errors observed with the human experts when recording ground truth data.

In addition to the inherent smoothness of the spline curve, another term penalizes large translations between control points. According to Euler–Bernoulli beam theory, the bending moment of the trajectory is proportional to its curvature, so we choose analogously

$$R_B(\mathbf{p}) = \sum_k \|\nabla_k^2 \mathbf{p}_k\|. \tag{6}$$

We iteratively estimate the optimal parameters for $E(\mathbf{p},\mathbf{q})$ by gradient descent. The updated parameters $\mathbf{p}_u$ and $\mathbf{q}_u$ are given by

$$\nabla E(\mathbf{p},\mathbf{q}) = \nabla S(\mathbf{p},\mathbf{q}) - \lambda_1 \cdot \nabla R_T(\mathbf{p},\mathbf{q}) - \lambda_2 \cdot \nabla R_B(\mathbf{p},\mathbf{q}) \tag{7}$$

$$(\mathbf{p}_u, \mathbf{q}_u) \leftarrow sign(\nabla E(\mathbf{p},\mathbf{q})) \cdot min(\tau \cdot \|\nabla E(\mathbf{p},\mathbf{q})\|, \delta), \tag{8}$$

where τ is a magnitude control parameter, and δ is the maximum step width.

A flow-chart of our method is presented in Fig. 2.

During optimization, only data for a neighborhood of frames needs to be available, so frames can be processed in sequence like with previous approaches. With real-time implementations, only frames up to the current time may be considered. However, with off-line applications all data is available, so the optimisation might "look ahead" for additional motion compensation.

Integration of other distance measures: The gradient is approximated via finite differences, and the framework is extensible to include other measures,

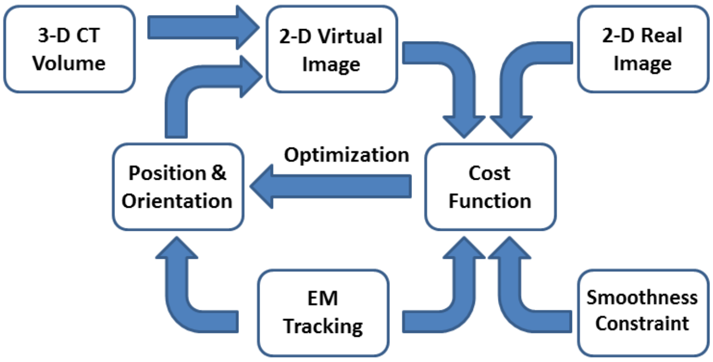

Fig. 2 Flowchart of hybrid EM and image-based tracking method. Position and orientation are initialized from EM tracking, and this initial estimate is used to create a virtual 2-D image from the 3-D CT volume. This virtual image is compared to the real image, and the image similarity, smoothness constraint and EM tracking data are combined in the cost function. In an optimization step, position and orientation are repeatedly updated, until a stop condition is fulfilled.

even when a closed form description of the gradient is not available. Equation 3 can be extended to use other measures, e.g. displacement information obtained from airway segmentation [38]. Such information can provide additional robustness, since outside airways image similarity is not meaningful. If optimization includes the distance from the nearest airway voxel, this can guide the optimization until image similarity is available again.

Registration Refinement: For each control point, its original EM tracking measurements and its corresponding corrected position and orientation are available. Thus, after a trajectory has been processed once, it is possible to compute the optimal Euclidean transform between both point sets, i.e. between original and corrected positions. This is similar to an update step of the iterative closest point (ICP) algorithm [5]. A visual inspection of all accessible branches is commonly recommended by medical guidelines prior to a bronchoscopic procedure [16], and the refined and approved registration can then be used for subsequent navigation and visualisation, well integrated into the clinical workflow.

3 Experiments

Evaluation Setup: We use a 3D Guidance EM tracking system (Ascension Technology, Burlington, VT, USA) with a flat-bed field generator and model 130 sensors (outer diameter 1.5 mm, sensor length 7.7 mm, 6-D), and a BF-P260F flexible fibre-optic bronchoscope (Olympus, Tokyo, Japan). One EM tracking sensor was fixed inside the bronchoscope working channel. The size of the endoscopic video frames is 362×370 pixels.

The dynamic phantom is a CLA 9 (CLA, Coburg, Germany), which contains five and more generations of bronchial branches and was chosen due to

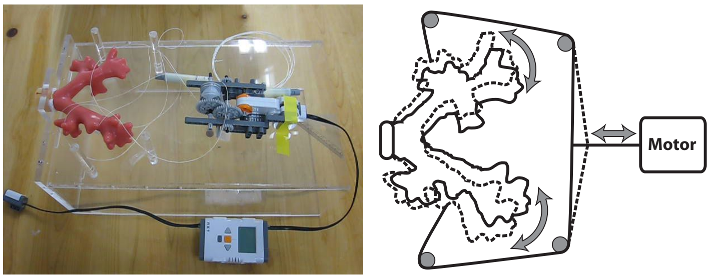

Fig. 3 Dynamic motion phantom (left) and description of phantom operation (right).

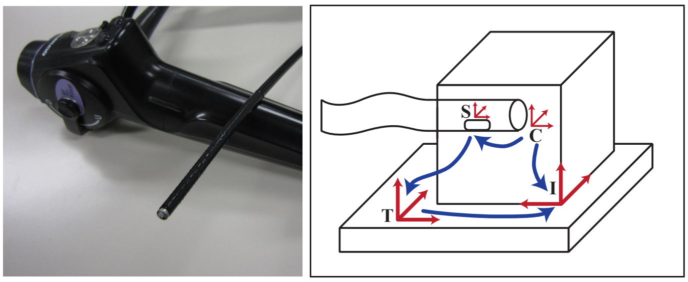

Fig. 4 Bronchoscope with embedded EM tracking sensor (left) and coordinate systems involved (right): bronchoscope camera (C), CT image (I), electromagnetic tracking (T), and tracking sensor (S) coordinate frames. The transformation ${}^{T}T_{S}$ is measured, the transformations ${}^{S}T_{C}$ and ${}^{I}T_{T}$ are calibrated, and the transformation ${}^{I}T_{C}$ is optimized for each frame.

its closely human-mimicking surface. It was connected to a motor (Lego, Billund, Denmark) via nylon threads (cf. Fig. 3), in order to physically separate the metallic motor components from the phantom, where the measurements were made. Four data sets consisting of video sequences and EM tracking data recordings were acquired with different amplitudes of simulated breathing motion between 7.48 and 23.65 mm, which corresponds to motion amplitudes determined for humans [15].

Both video sequence and CT data were stored in GPU memory, and virtual bronchoscopic image rendering as well as similarity were computed on GPU using OpenGL. Experiments were conducted on a standard workstation (Intel Core Duo 6600, NVidia GeForce 8800 GTS with 96 cores).

A CT scan of the phantom was acquired with 0.5 mm slice thickness, the resulting volume was $512 \times 512 \times 1021$ voxels. The bronchoscope and the different coordinate systems and transformations are shown in Fig. 4. For point-based registration between CT and EM tracking coordinate systems, 29 external landmarks were used, and average residual error was 0.92 mm.

Other methods for registration between image and tracking coordinates could be applied as well, e.g. using the ICP algorithm for matching recorded positions to the airway centerlines [8], optimizing the fraction of recorded po-

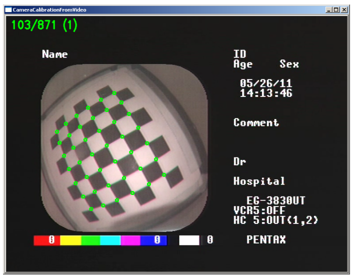

Fig. 5 Camera calibration screenshot with segmented inner corners of calibration pattern. Please note the strong barrel distortion, or "fisheye" effect.

sitions within the airways [23], or possibly even include a first-order estimation of breathing motion [11].

Camera calibration: Camera operation is described by the pin-hole camera model. A high-contrast pattern with known geometry is used, from which features are extracted, similar to Fig. 5. An estimate of the intrinsic parameters is computed first, using a closed-form solution, and then all parameters, including distortion, are optimised using an iterative technique as presented by Wengert *et al.* [50].

3.1 Ground Truth Data

In previous works, tracking of bronchoscopic video relative to CT data was usually evaluated via the duration of successful tracking [31], the fraction of "correctly matched frames" [28], displacement from airway centreline [49], fraction of positions within airways [23], or distance between successive result positions [13]. Such criteria are either strongly dependent on data sets and a subjective definition of "successful tracking", or biased if the very same criteria have been optimised before. The ground truth data problem was explained in details by Mountney *et al.* [33].

As an extension of the "manual registration" approach by Soper *et al.* [45], we proposed an evaluation based on expert-provided ground truth data [39]. A direct deduction of clinical relevance is possible, and from the agreement within experts and between multiple experts the limits of image-based methods can be learned.

Ground truth data is independently and repeatedly collected by two experts, one expert bronchoscopist (A) and one scientist (B). Recording is started from the first frame in each video sequence and a neutral, approximately correct position for the first frame. For each real image the position and orientation of the virtual bronchoscopic image are manually adjusted using mouse and keyboard, until both images match as closely as possible. Each expert is blinded to the other expert's results, as well as to his own results from previous sessions. For each data set, the procedure was repeated up to four times with each expert. Since this process is extremely time-consuming, only a subset of frames may be matched, evenly distributed over the full image sequence.

For analysis, pose data from multiple sessions is first averaged per expert, then these intermediate results are averaged between all experts. Intra-expert agreement and inter-expert agreement are computed as mean standard deviation of selected poses and averaged over all poses, either for each single expert's poses, or for all experts. These agreement values can then be used to indicate general limits for approaches based on registration of real and virtual bronchoscopy images.

In our case intra-expert agreement was 1.66 mm and 5.80° (A) and 1.44 mm and 3.94° (B), and inter-expert agreement was 1.26 mm and 4.78°. The ratio between intra- and inter-expert agreement also indicates considerable overlap between the experts' results.

3.2 Accuracy Computation

Accuracy is typically used as a synonym of "trueness", i.e. the expected deviation of a measurement from the true reference value:

$$\mu_p = \frac{1}{n}\sum_{i=1}^{n}\|\mathbf{p}_i - \hat{\mathbf{p}}_i\| \qquad \text{and} \qquad \mu_q = \frac{1}{n}\sum_{i=1}^{n}\theta(\mathbf{q}_i, \hat{\mathbf{q}}_i), \tag{9}$$

for translation and rotation, respectively, where $\mathbf{p}_i$ and $\mathbf{q}_i$ is the i-th measurement of a fiducial, and $\hat{\mathbf{p}}_i$ and $\hat{\mathbf{q}}_i$ is the fiducial's "forever unknown" [12] true position and orientation, for which we acquired expert estimates. Accuracy was evaluated for all frames, where ground truth was available. For frames not at control points, their interpolated position and orientation was compared to ground truth.

The mean inter-frame distance as a measure of smoothness was defined for a sequence of positions $\mathbf{p}_1 \ldots \mathbf{p}_n$ as

$$\frac{1}{n-1}\sum_{i=1}^{n-1}\|\mathbf{p}_i - \mathbf{p}_{i+1}\| \qquad \text{and} \qquad \frac{1}{n-1}\sum_{i=1}^{n-1}\theta(\mathbf{q}_i, \mathbf{q}_{i+1}), \tag{10}$$

for translation and rotation, respectively.

4 Results and Comparison

We compare the proposed method to our own implementations of four previously published approaches, which have already been applied to similar trajectories: bronchoscope tracking by EM tracking only [44], intensity-based registration (IBR) with direct initialization from EM tracking [31], IBR with dynamic initialisation from EM tracking [27], and IBR with a sequential Monte Carlo sampler based on EM tracking [28]. Quantitative results for the accuracy

	Accuracy	Smoothness
Solomon *et al.* [44]	5.87 ± 2.67 mm $10.55 \pm 6.16°$	3.20 ± 1.68 mm $3.40 \pm 9.22°$
Mori *et al.* [31]	5.59 ± 2.91 mm $10.55 \pm 6.37°$	3.72 ± 2.23 mm $4.98 \pm 10.89°$
Luo *et al.* [27]	5.17 ± 3.29 mm $10.75 \pm 6.79°$	3.24 ± 2.24 mm $4.57 \pm 10.57°$
Luo *et al.* [28]	4.52 ± 3.08 mm $10.62 \pm 6.14°$	2.77 ± 2.08 mm $3.46 \pm 9.46°$
Proposed Method	4.91 ± 2.57 mm $11.48 \pm 6.09°$	$\mathbf{1.24 \pm 0.82}$ **mm** $\mathbf{3.00 \pm 8.36°}$
Expert agreement/motion	1.26 mm 4.78°	1.21 mm 3.03°

Table 1 Tracking accuracy and smoothness of our method compared to previous approaches. In middle column, mean error and standard deviation from to ground truth are given for translation and rotation. Accuracy was evaluated for all frames, i.e. not only control points, and is equivalent within expert agreement. In right column, mean inter-frame distance and standard deviation are given for translation and rotation. Smoothness is significantly improved by the proposed method, compared to previous approaches, which do not enforce smoothness.

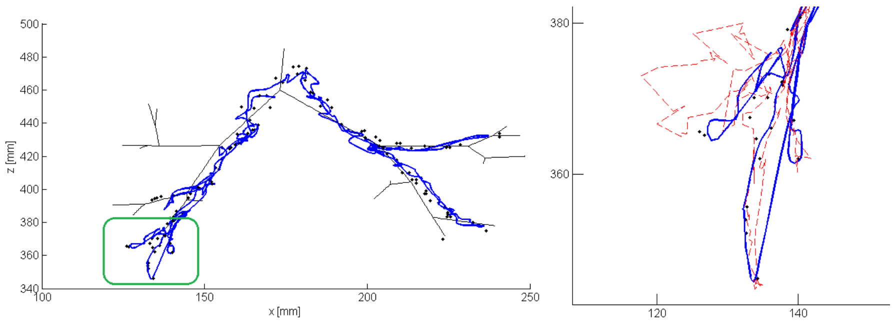

Fig. 6 Left: frontal view of airway structure, trajectory from proposed method (solid blue), and ground truth positions (black dots). Close-up region is indicated as green rectangle. Right: close-up view of trajectories from the previous approach [28] (dashed red) and our proposed method (solid blue), showing much smoother output of the latter.

relative to ground truth, and smoothness of the output in terms of inter-frame distances are given in Table 1.

Accuracy was evaluated relative to expert-provided data, and the accuracy of the proposed method is equivalent to previous approaches with regard to expert agreement. In addition, our results for the accuracy of previous methods agree with Soper *et al.* [45], who reported mean displacements of 2.37 mm and 8.46°.

The proposed method shows significantly better smoothness than other methods. In order to also visualize the difference, two output trajectories are shown in Fig. 6. The close-up view shows significantly less jitter in the output of the proposed method than of the previous approach previously presented by Luo *et al.* [28], which did not enforce smoothness. As shown in the close-up view, there were several local minima of the cost function in close proximity, which the optimization reached in succession. In contrast, due to enforcement of smoothness with the proposed method, these minima were less attractive during optimization.

It is important to note that the inter-frame distance cannot reach zero, since it needs to accommodate the true motion as well. An estimate for the true motion was computed as the average motion per frame between ground truth poses defined by the human experts.

Computation time is 0.98 seconds per frame, i.e. between 12.8 and 14.9 minutes for the 781 to 911 frames of our video sequences. For other methods a run time between 0.75 and 2.5 seconds [31,27,45,28] per frame is reported, so our computation speed is equivalent to previous approaches.

5 Discussion

The accuracy of our method is equivalent to previous approaches, with a significant improvement in smoothness. The remaining inter-frame distance is close to the real bronchoscope motion, as determined from ground truth data. Due to the smoothness improvement, our method by design provides a more realistic and physically plausible solution to the tracking problem, which has significantly less jitter. The quantitatively better smoothness results are confirmed by video output[1], which is much more consistent and robust, i.e. with fewer occasions of tracking loss or unexpected movement. In particular, the significant improvement of angular jitter can be beneficial for augmented reality set-ups. The filtering effect removes noise from the input, but can also lead to some loss of detail, if e.g. the direction of motion changes. Smoothness could also be achieved (at the cost of reducing accuracy and introducing latency) with filtering, but with medical applications any loss of accuracy would be hard to justify.

Soper *et al.* [45] reported a slightly lower error of 3.33 mm for their method of tracking an ultrathin bronchoscope. However, the actual clinical relevance of ultrathin bronchoscopes is low, as those have not yet proven diagnostic advantages over conventional bronchoscopes [47,53] and do not provide any steering mechanism. Also, Soper *et al.* only considered comparatively simple trajectories consisting of one smooth, one-way motion from the trachea into the peripheral lung. In contrast, recommendations for bronchoscopic procedures commonly list e.g. the need to visually inspect all accessible branches at the beginning of an intervention [16]. A typical trajectory will therefore consist of much more irregular motion and more changes of direction.

We expect our method to be robust against transient artifacts occurring in real clinical images, like specular reflections or bubbles, since groups of frames are matched and smoothness is enforced. We used a particularly large number of landmarks, in order to have an optimal initial registration for evaluation. In clinical routine there will also be fewer landmarks available: Typically five

[1] `http://campar.in.tum.de/files/publications/reichl2012ijcars.video.avi`

to seven bronchial bifurcations can be used for registration [40]. However, improvement of initial registration has been the focus of other works already [8, 23, 11].

Our dynamic motion phantom contains five and more generations of bronchial branches (cf. Fig. 6). During clinical use, a set of smaller branches might be encountered, but even when using ultrathin bronchoscopes, progress was rather limited by bronchus size than by image-based tracking [45]. In the current phantom design movement is mostly within the coronal plane and left-right. However, our method does not impose any assumptions on the breathing motion, which, depending on the degree of patient sedation, can be irregular and interrupted by coughing or choking.

Future work will definitely include acquisition of ground truth data from more interventional experts. However, in clinical routine time is a precious resource and the acquisition process is extremely time-consuming. When working accurately, approximately 30-40 poses can be recorded per hour. In our case, manual adjustment was done using mouse and keyboard, but in the future could be performed more ergonomically with 6-D input devices like the SpaceNavigator (3Dconnexion, München, Germany) or PHANTOM Omni (Sensable, Wilmington, MA, USA). We would not expect any effect in the accuracy of obtained results, since frames were optimized until the images matched satisfactorily and time was not limited. However, we would expect data collection to be performed quicker, when more advanced input devices are used.

As an alternative ground truth, landmarks could have been determined in the CT image and videos, but then only in a fraction of the video images those landmarks are visible, and especially the performance of a method in the absence of landmarks is relevant.

The next step will be applying the proposed method to trajectories recorded during actual interventions in humans. Since the dynamic motion phantom sequences were recorded anticipating the real procedure, data acquisition can be brought to the operating room with minimal interruption to the surgical workflow.

Currently, our method uses iso-surface volume rendering, which allows real-time adaption of rendering parameters, including the iso-value, which was beneficial during development. However, the surface could be pre-computed using airway segmentation, as proposed by e.g. Gergel *et al.* [13], and rendering speed could be improved in the future using surface rendering. In addition, current high-end GPUs could be used to take full advantage of the inherently possible parallel computation of the cost function.

Electromagnetic tracking in principle can be negatively affected by metallic components in the environment through Eddy currents and secondary EM fields. However, off-the shelf cytology needles have been equipped with EM sensors and found not to affect EM tracking [14]. Table-top field generators are commercially available for the 3D Guidance (Ascension Technology, Burlington, VT, USA) and Aurora (Northern Digital, Waterloo, ON, Canada) EM tracking systems, which are immune to distortions from metals below the transmitter, e.g. from operating room tables.

One weakness of bronchoscope tracking by intensity-based registration of images is that the position accuracy is worse along the view direction than laterally. Translations along the view direction result in only small changes in the image similarity, and this is obscured further by the remaining, unavoidable discrepancies between the pre-operative CT volume and the intra-operative video. In the future, tracking of flexible endoscopes might be improved by matching 3-D data to the 3-D CT volume. "Shape from shading" [37,35] has been applied to bronchoscope tracking [6,10,46], and registration of 3-D point clouds to 3-D surfaces from CT data has been proposed for image-based tracking [30]. Such approaches do not depend on a photo-realistic rendering sub-step, but the shape-from-shading sub-step in turn is affected by specularities and mucus on the airway surface.

6 Conclusions

We presented a novel approach to hybrid imaged-based and electromagnetic tracking of flexible bronchoscopes by modeling EM tracking output as spatially smooth over time. While providing equivalent accuracy at equivalent speeds, compared to previous methods we significantly improve inter-frame smoothness from 2.77–3.72 mm and 11.48° to 1.24 mm and and 3.00°. Due to the smoothness improvement, our method by design provides a more realistic and physically plausible solution to the tracking problem, which has significantly less jitter. We have outlined possibilities for future research and possible improvements. As we only impose minimal prior knowledge about the visual appearance of anatomy, but do not depend on e.g. airway segmentation, tree structure, or bronchoscope properties, our method can be applied to other endoscopic applications as well.

Acknowledgements: This work was supported by the Deutsche Forschungsgemeinschaft under grants NA 620/2-1 and 446 JAP 113/348/0-1, the European Union FP7 under grant n°256984, the JSPS Grant for Scientific Research, and the TUM Graduate School of Information Science in Health (GSISH). The authors would like to thank Marco Feuerstein for help with Fig. 6.

Conflict of interest: The authors declare that they have no conflict of interest.

Ethical standards: All human and animal studies have been approved by the appropriate ethics committee and have therefore been performed in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki and its later amendments. All persons gave their informed consent prior to their inclusion in the study.

References

1. Atasoy, S., Noonan, D.P., Benhimane, S., Navab, N., Yang, G.Z.: A global approach for automatic fibroscopic video mosaicing in minimally invasive diagnosis. In: Proceedings of Int'l Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), *Lecture Notes In Computer Science*, vol. 5241, pp. 850–857. Springer Berlin / Heidelberg, New York, USA (2008)
2. Azuma, R., Bishop, G.: Improving static and dynamic registration in an optical see-through hmd. In: Proceedings of the 21st Annual Conference on Computer Graphics and Interactive Techniques, pp. 197–204. ACM (1994)
3. Baillot, Y., Eliason, J.J., Schmidt, G.S., Swan J. E., I., Brown, D., Julier, S., Livingston, M.A., Rosenblum, L.: Evaluation of the ShapeTape tracker for wearable, mobile interaction. In: Proc. IEEE Virtual Reality, pp. 285–286 (2003)
4. Balter, J.M., Wright, J.N., Newell, L.J., Friemel, B., Dimmer, S., Cheng, Y., Wong, J., Vertatschitsch, E., Mate, T.P.: Accuracy of a wireless localization system for radiotherapy. International Journal of Radiation Oncology*Biology*Physics **61**(3), 933 – 937 (2005). DOI 10.1016/j.ijrobp.2004.11.009. URL `http://www.sciencedirect.com/science/article/pii/S0360301604028391`
5. Besl, P., McKay, H.: A method for registration of 3-D shapes. IEEE Transactions on Pattern Analysis and Machine Intelligence **14**(2), 239–256 (1992)
6. Bricault, I., Ferretti, G., Cinquin, P.: Registration of real and CT-derived virtual bronchoscopic images to assist transbronchial biopsy. IEEE Transactions on Medical Imaging **17**(5), 703–714 (1998)
7. Catmull, E., Rom, R.: A class of interpolating splines. Computer Aided Geometric Design pp. 317–326 (1974)
8. Deguchi, D., Ishitani, K., Kitasaka, T., Mori, K., Suenaga, Y., Takabatake, H., Mori, M., Natori, H.: A method for bronchoscope tracking using position sensor without fiducial markers. In: Proceedings of SPIE Medical Imaging, vol. 6511, p. 65110N. SPIE (2007)
9. Deguchi, D., Mori, K., Feuerstein, M., Kitasaka, T., Maurer, C.R., Suenaga, Y., Takabatake, H., Mori, M., Natori, H.: Selective image similarity measure for bronchoscope tracking based on image registration. Medical Image Analysis **13**(4), 621–633 (2009)
10. Deligianni, F., Chung, A., Yang, G.Z.: Patient-specific bronchoscope simulation with pq-space-based 2D/3D registration. Computer Aided Surgery **9**(5), 215–226 (2004)
11. Feuerstein, M., Sugiura, T., Deguchi, D., Reichl, T., Kitasaka, T., Mori, K.: Marker-free registration for electromagnetic navigation bronchoscopy under respiratory motion. In: Proc. Int'l Workshop on Medical Imaging and Augmented Reality (MIAR), *Lecture Notes In Computer Science*, vol. 6326, pp. 237–246. Springer Berlin / Heidelberg (2010)
12. Fitzpatrick, J.M., West, J.B., Maurer Jr., C.R.: Predicting error in rigid-body point-based registration. IEEE Transactions on Medical Imaging **14**(5), 694–702 (1998)
13. Gergel, I., dos Santos, T.R., Tetzlaff, R., Maier-Hein, L., Meinzer, H.P., Wegner, I.: Particle filtering for respiratory motion compensation during navigated bronchoscopy. In: K.H. Wong, M.I. Miga (eds.) Proceedings of SPIE Medical Imaging 2010: Visualization, Image-Guided Procedures, and Modeling, vol. 7625, p. 76250W. SPIE (2010)
14. Gergel, I., Tetzlaff, R., Meinzer, H.P., Wegner, I.: Evaluation of electromagnetically tracked transbronchial needle aspiration in a ventilated porcine lung. In: Proceedings of SPIE Medical Imaging 2011: Visualization, Image-Guided Procedures, and Modeling, pp. 79,640P–79,640P–6 (2011)
15. Hanley, J., Debois, M.M., Mah, D., Mageras, G.S., Raben, A., Rosenzweig, K., Mychalczak, B., Schwartz, L.H., Gloeggler, P.J., Lutz, W., Ling, C.C., Leibel, S.A., Fuks, Z., Kutcher, G.J.: Deep inspiration breath-hold technique for lung tumors: the potential value of target immobilization and reduced lung density in dose escalation. International Journal of Radiation Oncology*Biology*Physics **45**(3), 603–611 (1999)
16. Häussinger, K., Ballin, A., Becker, H.D., Bölcskei, P., Dierkesmann, R., Dittrich, I., Frank, W., Freitag, L., Gottschall, R., Guschall, W.R., Hartmann, W., Hauck, R., Herth, F., Kirsten, D., Kohlhäufl, M., Kreuzer, A., Loddenkemper, R., Macha, N., Markus, A., Stanzel, F., Steffen, H., Wagner, M.: Recommendations for quality standards in bronchoscopy. Pneumologie **58**(5), 344–356 (2004)

17. Hautmann, H., Schneider, A., Pinkau, T., Peltz, F., Feussner, H.: Electromagnetic catheter navigation during bronchoscopy: validation of a novel method by conventional fluoroscopy. Chest **128**(1), 382–387 (2005)
18. Hekimian-Williams, C., Grant, B., Liu, X., Zhang, Z., Kumar, P.: Accurate localization of RFID tags using phase difference. In: RFID, 2010 IEEE International Conference on, pp. 89 –96 (2010)
19. Höller, K., Penne, J., Schneider, A., Jahn, J., Boronat, J.G., Wittenberg, T., Feussner, H., Hornegger, J.: Endoscopic orientation correction. In: Proceedings of Int'l Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), *Lecture Notes In Computer Science*, vol. 5761, pp. 459–466. Springer Berlin / Heidelberg (2009)
20. Housden, R., Treece, G., Gee, A., Prager, R., Street, T.: Hybrid systems for reconstruction of freehand 3D ultrasound data. CUED/F-INFENG/TR 574, University of Cambridge, Department of Engineering (2007)
21. Hummel, J., Figl, M., Bax, M., Shahidi, R., Bergmann, H., Birkfellner, W.: Evaluation of dynamic electromagnetic tracking deviation. In: M.I. Miga, K.H. Wong (eds.) Proceedings of SPIE Medical Imaging 2009: Visualization, Image-Guided Procedures, and Modeling, vol. 7261, p. 72612U. SPIE (2009)
22. Jemal, A., Bray, F., Center, M.M., Ferlay, J., Ward, E., Forman, D.: Global cancer statistics. CA Cancer Journal for Clinicians **61**(2), 69–90 (2011)
23. Klein, T., Traub, J., Hautmann, H., Ahmadian, A., Navab, N.: Fiducial-free registration procedure for navigated bronchoscopy. In: Proceedings of Int'l Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), *Lecture Notes In Computer Science*, vol. 4791, p. 475. Springer Berlin / Heidelberg (2007)
24. Koizumi, N., Sumiyama, K., Suzuki, N., Hattori, A., Tajiri, H., Uchiyama, A.: Development of three-dimensional endoscopic ultrasound system with optical tracking. In: T. Dohi, R. Kikinis (eds.) Proceedings of Int'l Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), *Lecture Notes In Computer Science*, vol. 2488, pp. 60–65. Springer Berlin / Heidelberg (2002)
25. Kupelian, P., Willoughby, T., Mahadevan, A., Djemil, T., Weinstein, G., Jani, S., Enke, C., Solberg, T., Flores, N., Liu, D., Beyer, D., Levine, L.: 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). DOI 10.1016/j.ijrobp.2006.10.026. URL `http://dx.doi.org/10.1016/j.ijrobp.2006.10.026`
26. Li, Y., Aissaoui, R., Lacoste, M., Dansereau, J.: Development and evaluation of a new body-seat interface shape measurement system. IEEE Transactions on Biomedical Engineering **51**(11), 2040–2050 (2004)
27. Luo, X., Feuerstein, M., Sugiura, T., Kitasaka, T., Imaizumi, K., Hasegawa, Y., Mori, K.: Towards hybrid bronchoscope tracking under respiratory motion: evaluation on a dynamic motion phantom. In: K.H. Wong, M.I. Miga (eds.) Proceedings of SPIE Medical Imaging 2010: Visualization, Image-Guided Procedures, and Modeling, vol. 7625, p. 76251B. SPIE (2010)
28. Luo, X., Reichl, T., Feuerstein, M., Kitasaka, T., Mori, K.: Modified hybrid bronchoscope tracking based on sequential Monte Carlo sampler: dynamic phantom validation. In: Asian Conference on Computer Vision, pp. 409–421. Queenstown, New Zealand (2010)
29. Markelj, P., Tomazevic, D., Likar, B., Pernus, F.: A review of 3D/2D registration methods for image-guided interventions. Medical Image Analysis **16**(3), 642–661 (2010)
30. Mirota, D., Taylor, R.H., Ishii, M., Hager, G.D.: Direct endoscopic video registration for sinus surgery. In: M.I. Miga, K.H. Wong (eds.) Proceedings of SPIE Medical Imaging 2009: Visualization, Image-Guided Procedures, and Modeling, vol. 7261, p. 72612K. SPIE (2009)
31. Mori, K., Deguchi, D., Akiyama, K., Kitasaka, T., Maurer, C.R., Suenaga, Y., Takabatake, H., Mori, M., Natori, H.: Hybrid bronchoscope tracking using a magnetic tracking sensor and image registration. In: Proceedings of Int'l Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), *Lecture Notes In Computer Science*, vol. 3750, pp. 543–550. Springer Berlin / Heidelberg (2005)

32. Mori, K., Hasegawa, J., Toriwaki, J., Anno, H., Katada, K.: Automated extraction and visualization of bronchus from 3D CT images of lung. In: Proceedings of the First International Conference on Computer Vision, Virtual Reality and Robotics in Medicine, *Lecture Notes In Computer Science*, vol. 905, pp. 542–548. Springer Berlin / Heidelberg (1995)
33. Mountney, P., Stoyanov, D., Yang, G.Z.: Three-dimensional tissue deformation recovery and tracking. Signal Processing Magazine, IEEE **27**(4), 14 –24 (2010)
34. Nakamoto, M., Ukimura, O., Gill, I., Mahadevan, A., Miki, T., Hashizume, M., Sato, Y.: Realtime organ tracking for endoscopic augmented reality visualization using miniature wireless magnetic tracker. In: T. Dohi, I. Sakuma, H. Liao (eds.) Proc. Int'l Workshop on Medical Imaging and Augmented Reality (MIAR), *Lecture Notes In Computer Science*, vol. 5128, pp. 359–366. Springer Berlin / Heidelberg (2008). DOI 10.1007/978-3-540-79982-5_39. URL `http://www.springerlink.com/content/j6846v117v1xw683/`
35. Okatani, T., Deguchi, K.: Shape reconstruction from an endoscope image by shape from shading technique for a point light source at the projection center. Computer Vision and Image Understanding **66**(2), 131 (1997)
36. Pagoulatos, N., Rohling, R.N., Edwards, W.S., Kim, Y.: New spatial localizer based on fiber optics with applications in 3D ultrasound imaging. In: S.K. Mun (ed.) Proceedings of SPIE Medical Imaging 2000: Image Display and Visualization, vol. 3976, pp. 595–602. SPIE (2000)
37. Rashid, H., Burger, P.: Differential algorithm for the determination of shape from shading using a point light source. Image and Vision Computing **10**(2), 119–127 (1992)
38. Reichl, T., Gergel, I., Menzel, M., Hautmann, H., Wegner, I., Meinzer, H.P., Navab, N.: Real-time motion compensation for EM bronchoscope tracking with smooth output – ex-vivo validation. In: Proceedings of SPIE Medical Imaging 2012: Image-Guided Procedures, Robotic Interventions, and Modeling, vol. 8316, p. 83163A. SPIE (2012)
39. Reichl, T., Luo, X., Menzel, M., Hautmann, H., Mori, K., Navab, N.: Deformable registration of bronchoscopic video sequences to CT volumes with guaranteed smooth output. In: Proceedings of Int'l Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), *Lecture Notes In Computer Science*, vol. 6891, pp. 17–24. Springer Berlin / Heidelberg, Toronto, Canada (2011)
40. Schwarz, Y., Greif, J., Becker, H.D., Ernst, A., Mehta, A.: Real-time electromagnetic navigation bronchoscopy to peripheral lung lesions using overlaid CT images: the first human study. Chest **129**(4), 988–994 (2006)
41. Shimada, J., Itoh, K., Yamaguchi, S., Nishikawa, A., Miyazaki, F.: Magnetic navigation system for thoracoscopic surgery of the pig lung partial resection with transbronchial marking. In: Proceedings of Int'l Congress on Computer Assisted Radiology and Surgery (CARS), vol. 1281, pp. 752–755. Elsevier (2005). DOI http://dx.doi.org/10.1016/j.ics.2005.03.164. URL `http://linkinghub.elsevier.com/retrieve/pii/S053151310500422X`
42. Shoemake, K.: Animating rotation with quaternion curves. In: Proceedings of Conference on Computer Graphics and Interactive Techniques (SIGGRAPH), pp. 245–254. ACM New York, NY, USA (1985)
43. Sluimer, I., Schilham, A., Prokop, M., van Ginneken, B.: Computer analysis of computed tomography scans of the lung: a survey. IEEE Transactions on Medical Imaging **25**(4), 385–405 (2006)
44. Solomon, S.B., White, P., Wiener, C.M., Orens, J.B., Wang, K.P.: Three-dimensional CT-guided bronchoscopy with a real-time electromagnetic position sensor: a comparison of two image registration methods. Chest **118**(6), 1783–1787 (2000)
45. Soper, T., Haynor, D., Glenny, R., Seibel, E.: In vivo validation of a hybrid tracking system for navigation of an ultrathin bronchoscope within peripheral airways. IEEE Transactions on Biomedical Engineering **57**(3), 736–745 (2010)
46. Suter, M.J., Reinhardt, J.M., Sonka, M., Higgins, W.E., Hoffman, E.A., McLennan, G.: Three-dimensional true color topographical analysis of the pulmonary airways. In: A.A. Amini, A. Manduca (eds.) Proceedings of SPIE Medical Imaging 2004: Physiology, Function, and Structure from Medical Images, vol. 5369, pp. 189–198. SPIE (2004)

47. Tachihara, M., Ishida, T., Kanazawa, K., Sugawara, A., Watanabe, K., Uekita, K., Moriya, H., Yamazaki, K., Asano, F., Munakata, M.: A virtual bronchoscopic navigation system under X-ray fluoroscopy for transbronchial diagnosis of small peripheral pulmonary lesions. Lung Cancer **57**(3), 322–327 (2007)
48. Tschirren, J., Hoffman, E.A., McLennan, G., Sonka, M.: Intrathoracic airway trees: segmentation and airway morphology analysis from low-dose CT scans. IEEE Transactions on Medical Imaging **24**(12), 1529–1539 (2005)
49. Wegner, I., Tetzlaff, R., Biederer, J., Wolf, I., Meinzer, H.: An evaluation environment for respiratory motion compensation in navigated bronchoscopy. In: Proceedings of SPIE Medical Imaging, vol. 6918, p. 691811. SPIE (2008)
50. Wengert, C., Reeff, M., Cattin, P., Szekely, G.: Fully automatic endoscope calibration for intraoperative use. In: Proceedings of Bildverarbeitung für die Medizin (BVM), Informatik aktuell, pp. 419–423. Springer Berlin / Heidelberg (2006)
51. Wille, A., Broll, M., Winter, S.: Phase difference based RFID navigation for medical applications. In: Proceedings of IEEE International Conference on RFID (RFID), pp. 98 –105 (2011)
52. Willoughby, T.R., Kupelian, P.A., Pouliot, J., Shinohara, K., Aubin, M., Roach, M., Skrumeda, L.L., Balter, J.M., Litzenberg, D.W., Hadley, S.W., Wei, J.T., Sandler, H.M.: Target localization and real-time tracking using the calypso 4d localization system in patients with localized prostate cancer. International Journal of Radiation Oncology*Biology*Physics **65**(2), 528–534 (2006). DOI 10.1016/j.ijrobp.2006.01.050. URL `http://dx.doi.org/10.1016/j.ijrobp.2006.01.050`
53. Yamamoto, S., Ueno, K., Imamura, F., Matsuoka, H., Nagatomo, I., Omiya, Y., Yoshimura, M., Kusunoki, Y.: Usefulness of ultrathin bronchoscopy in diagnosis of lung cancer. Lung Cancer **46**(1), 43–48 (2004)