

Deformable Registration of Bronchoscopic Video Sequences to CT Volumes with Guaranteed Smooth Output

Tobias Reichl¹, Xiongbiao Luo², Manuela Menzel³, Hubert Hautmann³,
Kensaku Mori⁴, and Nassir Navab¹

¹ Computer-Aided Medical Procedures, TUM, München, Germany

² Graduate School of Information Science, Nagoya University, Japan

³ Medizinische Klinik I, Klinikum rechts der Isar, TUM, München, Germany

⁴ Information and Communications Headquarters, Nagoya University, Japan

Abstract. We present a novel approach to tracking of flexible bronchoscopes by modeling the output as spatially continuous over time. Bronchoscopy is a widespread clinical procedure for diagnosis and treatment of lung diseases and navigation systems are highly needed. Tracking of the bronchoscope can be regarded as a deformable registration problem. In our approach we use hybrid image-based and electromagnetic tracking, and the bronchoscope pose relative to CT data is interpolated using Catmull-Rom splines for position and SLERP for orientation.

We evaluate the method using ground truth poses manually selected by experts, where mean inter-expert agreement was determined as 1.26 mm. For four dynamic phantom data sets, the accuracy of our method is between 4.13 and 5.93 mm and shown to be equivalent to previous methods. We significantly improve inter-frame smoothness from 2.35–3.08 mm to 1.08–1.51 mm. Our method provides a more realistic and physically plausible solution with significantly less jitter. This quantitative result is confirmed by video output, which is much more consistent and robust, with fewer occasions of tracking loss or unexpected movement.

1 Introduction

Flexible bronchoscopy is a widespread clinical procedure for diagnosis and treatment of lung diseases, and one of its most common applications is transbronchial biopsy. This is commonly performed after lesions have been identified on Computed Tomography (CT) images, so it will definitely be beneficial to transfer this 3D information to the operating room. Since bronchoscopy is an inherently monitor-based procedure, augmentation of video images with guidance or targeting information is straightforward and promises high clinical acceptance due to smooth integration into the clinical workflow. The combination of flexible bronchoscopy with electromagnetic tracking (EMT) was first reported by Solomon et al. [1], and hybrid image-based and EM tracking was proposed by Mori et al. [2], and was improved by Luo et al. [3, 4] and Soper et al. [5]. Hybrid tracking

complements the advantages of image-based tracking and EMT, since image-based tracking is inaccurate for frames with little to no information due to low contrast. EMT suffers from noise and artifacts while the sensor is moved [6], and relative errors when anatomy is moving due to respiration or heartbeat.

While accuracy and computational speed are clearly important for real-time applications like intraoperative navigation, smoothness of the output is no less relevant. Less jitter of augmented reality overlays increases operator comfort and acceptance. Smooth output may provide benefits when further processing the output like for motion models, or temporal and spatial synchronization of multiple video sequences in the case of repeat examinations or clinical studies.

Tracking of the bronchoscope can be seen as a deformable registration problem as shown below, and in this work we introduce a novel solution to this problem. The deformable registration of the images over the bronchoscope trajectory is performed using Catmull-Rom splines and spherical linear interpolation (SLERP). Smoothness between consecutive frames is ensured using interpolation. We present a thorough and quantitative evaluation with respect to expert-provided ground truth data, including the determination of intra- and inter-expert agreement, and a quantitative comparison to the state of art.

2 Methods

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.

Discrete control points for the trajectory are set equidistant over time with spacing s . For interpolation of position we use Catmull-Rom splines [7], because the resulting curve is continuously differentiable and passes directly through the control points:

$$p(t) = \frac{1}{2} \begin{pmatrix} 1 & u & u^2 & u^3 \end{pmatrix} \begin{pmatrix} 0 & 2 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ 2 & -5 & 4 & -1 \\ -1 & 3 & -3 & 1 \end{pmatrix} \begin{pmatrix} p_{i-1} \\ p_i \\ p_{i+1} \\ p_{i+2} \end{pmatrix}, \quad (1)$$

where $p_{i-1} \dots p_{i+2}$ are positions of consecutive control points, $i = \lfloor t/s \rfloor$ is a control point index, and $u = t/s - \lfloor t/s \rfloor$ is the interpolation ratio between control points p_i and p_{i+1} .

For orientation we use quaternions, because they allow a continuous representation. Then, for interpolation between quaternions we use SLERP [8]:

$$q(t) = q_i (q_i^{-1} q_{i+1})^u = \frac{\sin(1-u)\theta}{\sin\theta} q_i + \frac{\sin u\theta}{\sin\theta} q_{i+1}, \quad (2)$$

where θ is the rotation difference between q_i and q_{i+1} . Initial parameters p_k and q_k for all control points are taken directly from the EMT measurements.

For the matching between bronchoscope images and CT image, we employ as an intermediate step virtual bronchoscopic images $I_V(p_k, q_k)$, where the CT image is rendered for a camera with position p_k and orientation q_k . We seek to maximize

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

where λ_1 and λ_2 are weighting parameters. The similarity $S(p, q)$ between real and virtual images is computed via the Local Normalized Cross Correlation (LNCC)

$$S(p, q) = \sum_k \text{LNCC}(I_R(t_k), I_V(p_k, q_k)), \quad (4)$$

where k is an image index, $I_R(t_k)$ is the real image at time t_k , and $I_V(p_k, q_k)$ is the virtual bronchoscopic image. The patch size for LNCC was set to 11. The gradient of $S(p, q)$ for movement for each single control point i can be approximated from the frames in its support $(i - 2, i + 2)$ via finite differences. Frames outside its support are not influenced by movement of this control point, so gradient computation 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.

According to Hooke's law, spring force is proportional to displacement, so we model tension, the cost of distance from the EM tracking measurements, as

$$R_T(p, q) = \sum_k \|p_k - p_{k,0}\| + \alpha \cdot \theta(q_k, q_{k,0}), \quad (5)$$

where $p_{k,0}$ and $q_{k,0}$ are the position and orientation measured by EM tracking at time t_k , and the rotation difference $\theta(\cdot)$ is computed as the angular component of the difference quaternion. α is the ratio between rotational and translational spring constants and was set to $3 \cdot 10^\circ / mm$, since this was 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

$$R_B(p) = \sum_k \|\nabla_k^2 p_k\|. \quad (6)$$

We iteratively estimate the optimal parameters for $E(p, q)$ by gradient descent. The update p_u, q_u is given by

$$\nabla E(p, q) = \nabla S(p, q) - \lambda_1 \cdot \nabla R_T(p, q) - \lambda_2 \cdot \nabla R_B(p, q) \quad (7)$$

$$(p_u, q_u) \leftarrow \text{sign}(\nabla E(p, q)) \cdot \min(\tau \cdot \|\nabla E(p, q)\|, \delta), \quad (8)$$

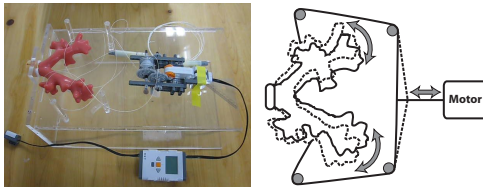


Fig. 1. Dynamic motion phantom (left) and description of phantom operation (right).

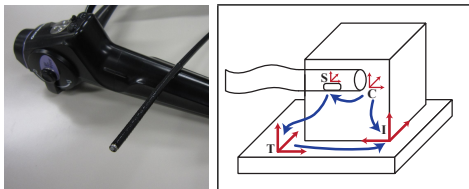


Fig. 2. Bronchoscope with embedded EMT sensor (left) and coordinate systems involved (right): bronchoscope camera (C), CT image (I), electromagnetic tracking (T), and tracking sensor (S) coordinate frames. The transformation ${}^T T_S$ is measured, the transformations ${}^S T_C$ and ${}^I T_T$ are calibrated, and the transformation ${}^I T_C$ is optimized for each frame.

where τ is a magnitude control parameter, and δ is the maximum step width. During optimization, only data for a neighborhood of frames needs to be available, so frames can be processed in sequence like with previous approaches.

3 Evaluation and Results

Setup: We use a 3D Guidance EMT system (Ascension, Burlington, USA) and a BF-P260F flexible fiberoptic bronchoscope (Olympus, Tokyo, Japan). One EMT sensor was fixed inside the bronchoscope working channel.

The dynamic phantom is a CLA 9 (CLA, Coburg, Germany), which was chosen due to its closely human-mimicking surface. It was connected to a motor (Lego, Billund, Denmark) via nylon threads (cf. Fig. 1). Four data sets consisting of video sequences and EMT data recordings were acquired with different amplitudes of simulated breathing motion between 7.48 and 23.65 mm.

A CT scan of the phantom was acquired with 0.5 mm slice thickness. The bronchoscope and the different coordinate systems and transformations are shown in Fig. 2. For point-based registration between CT and EMT coordinate systems, 29 external landmarks were used, average residual error was 0.92 mm. Camera intrinsics, deformation, and hand-eye calibration were performed using established methods [9].

Both video sequence and CT data were stored in graphics processing unit (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).

As an extension of the “manual registration” approach by Soper et al. [5] we chose an evaluation based on expert-provided ground truth data, since then a direct deduction of clinical relevance is possible, and from multiple experts the limits of image-based methods might be learned.

Ground truth data was independently and repeatedly collected by two experts, one expert bronchoscopist (A) and one scientist (B). For each real image, the position and orientation of the virtual bronchoscopic image were manually adapted via mouse and keyboard, until they matched the real image as closely as possible. Each expert was blinded to the other expert’s results, as well as to his own results from previous sessions. Due to the time-consuming process, only approximately every eighth frame was matched. First, pose data from multiple sessions was averaged per expert, then pose data from both experts.

Intra-expert agreement (mean standard deviation) was 1.66 mm and 5.80° (A) and 1.44 mm and 3.94° (B). Inter-expert agreement was 1.26 mm and 4.78° . The ratio between intra- and inter-expert agreement indicates considerable overlap between the experts’ results. These margins might indicate a limit for any approach based on registration of real and virtual bronchoscopy images.

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 EMT only [1], intensity-based registration (IBR) with direct initialization from EMT [2], IBR with dynamic initialization from EMT [3], and IBR with a Sequential Monte Carlo sampler based on EMT [4]. Quantitative results for the accuracy are given in Table 1. The precision of our evaluation method is defined by expert agreement, and within those bounds the accuracy of the proposed method is equivalent to previous approaches.

In Table 2 quantitative results for the smoothness of the output are given in terms of inter-frame distances. The results for previous methods agree with Soper et al. [5], who reported mean distances of 2.37 mm and 8.46° . The proposed method shows significantly better smoothness. In order to also visualize the difference, two output trajectories are shown in Fig. 3. In particular, the close-up view shows significantly less jitter in the output of the proposed method than of the previous approach.

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 [2, 3, 5, 4] per frame is reported, so our computation speed is equivalent to previous approaches.

4 Discussion

The accuracy of our method is equivalent to previous approaches, with a significant improvement in smoothness. The actual limit to smoothness is unknown, but we attribute most of the remaining inter-frame distance to the real bronchoscope motion. Due to the smoothness constraint, our method by design provides a more realistic and physically plausible solution to the tracking problem, which has significantly less jitter. This quantitative result is confirmed by video out-

Accuracy	Data set 1 n=116	Data set 2 n = 113	Data set 3 n = 110	Data set 4 n = 107
Solomon et al. [1]	4.90 ± 2.41 mm 8.99 ± 5.78°	5.86 ± 2.53 mm 9.50 ± 4.13°	5.71 ± 2.53 mm 10.60 ± 4.59°	7.00 ± 3.22 mm 13.09 ± 10.16°
Mori et al. [2]	4.57 ± 2.57 mm 9.13 ± 5.86°	5.76 ± 2.65 mm 9.52 ± 4.34°	5.34 ± 2.95 mm 10.32 ± 4.86°	6.69 ± 3.46 mm 13.24 ± 10.43°
Luo et al. [3]	4.12 ± 2.65 mm 9.29 ± 5.69°	5.11 ± 2.76 mm 9.10 ± 4.48°	4.77 ± 2.98 mm 10.29 ± 5.18°	6.68 ± 4.78 mm 14.33 ± 11.82°
Luo et al. [4]	3.95 ± 2.78 mm 9.13 ± 5.76°	4.00 ± 2.65 mm 9.58 ± 4.17°	4.80 ± 3.28 mm 10.65 ± 4.50°	5.32 ± 3.58 mm 13.13 ± 10.13°
Proposed Method	4.13 ± 2.30 mm 9.79 ± 4.99°	4.65 ± 2.58 mm 10.81 ± 5.22°	4.93 ± 2.41 mm 11.08 ± 4.67°	5.93 ± 2.97 mm 14.26 ± 9.49°
Expert agreement	1.27 mm 5.21°	1.15 mm 3.64°	1.33 mm 4.63°	1.27 mm 5.23°

Table 1. Accuracy comparison of approaches. In each cell, mean error and standard deviation with respect to ground truth are given for translation as well as rotation. Expert agreement denotes inter-expert standard deviation for each particular dataset. The results are equivalent with regard to expert agreement.

put⁵, which is much more consistent and robust, with fewer occasions of tracking loss or unexpected movement. 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.

Soper et al. [5] report 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 [10, 11] 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 [12]. 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.

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. As an alter-

⁵ <http://campar.in.tum.de/files/publications/reichl2011miccai.video.avi>

Smoothness	Data set 1	Data set 2	Data set 3	Data set 4
Solomon et al. [1]	2.86 ± 1.62 mm $2.11 \pm 8.21^\circ$	3.47 ± 1.72 mm $3.84 \pm 8.59^\circ$	3.01 ± 1.65 mm $3.81 \pm 11.25^\circ$	3.46 ± 1.73 mm $3.83 \pm 8.81^\circ$
Mori et al. [2]	3.35 ± 2.09 mm $3.50 \pm 9.68^\circ$	3.87 ± 2.31 mm $4.74 \pm 8.63^\circ$	3.61 ± 2.18 mm $6.05 \pm 14.51^\circ$	4.04 ± 2.35 mm $5.63 \pm 10.74^\circ$
Luo et al. [3]	2.86 ± 2.11 mm $3.22 \pm 10.42^\circ$	3.43 ± 2.30 mm $4.54 \pm 8.83^\circ$	3.11 ± 2.20 mm $5.17 \pm 12.35^\circ$	3.54 ± 2.35 mm $5.36 \pm 10.69^\circ$
Luo et al. [4]	2.35 ± 1.88 mm $2.12 \pm 8.24^\circ$	2.95 ± 2.09 mm $3.94 \pm 8.92^\circ$	2.71 ± 2.08 mm $3.84 \pm 11.32^\circ$	3.08 ± 2.28 mm $3.96 \pm 9.35^\circ$
Proposed Method	1.23 ± 0.76 mm $1.96 \pm 5.82^\circ$	1.51 ± 1.04 mm $3.18 \pm 6.23^\circ$	1.12 ± 0.69 mm $3.22 \pm 11.71^\circ$	1.08 ± 0.77 mm $3.62 \pm 9.66^\circ$

Table 2. Smoothness comparison of approaches. In each cell, mean inter-frame distance and standard deviation are given for translation as well as rotation. Smoothness is significantly improved by the proposed method.

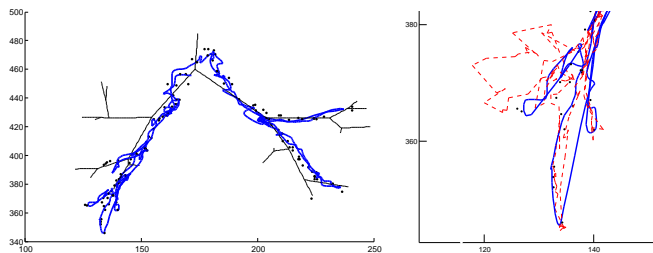


Fig. 3. Left: frontal view of airway structure, trajectory from proposed method (solid blue), and ground truth positions (black dots). Right: close-up view of trajectories from previous approach [4] (dashed red) and proposed method (solid blue), showing much smoother output of the latter.

native 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.

5 Conclusions

We present a novel approach to hybrid imaged-based and electromagnetic tracking of flexible bronchoscopes by modeling the output as spatially smooth over time. While providing equivalent accuracy at equivalent speeds, we significantly

improve smoothness. 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 Deutsche Forschungsgemeinschaft under grants NA 620/2-1 and 446 JAP 113/348/0-1, European Union FP7 under grant n°256984, JSPS Grant for Scientific Research, and the TUM Graduate School of Information Science in Health (GSISH). Thanks to M. Feuerstein for help with Fig. 3.

References

1. Solomon, S.B., White Jr., 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** (2000) 1783–1787
2. 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 Duncan, J., Gerig, G., eds.: MICCAI. Volume 3750 of LNCS. (2005) 543–550
3. 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: Proc. SPIE. Volume 7625. (2010) 76251B
4. Luo, X., Reichl, T., Feuerstein, M., Kitasaka, T., Mori, K.: Modified hybrid bronchoscope tracking based on sequential monte carlo sampler: Dynamic phantom validation. In: ACCV. (2010)
5. 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** (2010) 736–745
6. Hummel, J., Figl, M., Bax, M., Shahidi, R., Bergmann, H., Birkfellner, W.: Evaluation of dynamic electromagnetic tracking deviation. In: Proc. SPIE. Volume 7261. (2009) 72612U
7. Catmull, E., Rom, R.: A class of interpolating splines. *Computer Aided Geometric Design* (1974) 317–326
8. Shoemake, K.: Animating rotation with quaternion curves. In: SIGGRAPH. (1985)
9. Wengert, C., Reeff, M., Cattin, P., Szekely, G.: Fully automatic endoscope calibration for intraoperative use. In: Bildverarbeitung in der Medizin. (2006)
10. 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** (2007) 322–327
11. 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** (2004) 43–48
12. Häussinger, K., Ballin, A., Becker, H., et al.: Recommendations for quality standards in bronchoscopy. *Pneumologie* **58** (2004) 344–356