Enhanced Differential Evolution to Combine Optical Mouse Sensor with Image Structural Patches for Robust Endoscopic Navigation

Xiongbiao Luo[1,2], Uditha L. Jayarathne[2],
A. Jonathan McLeod[2], and Kensaku Mori[1]

[1] Information and Communications Headquarters, Nagoya University, Japan
[2] Robarts Research Institute, Western University, Canada
{xluo,ujayarat,jmcleod}@robarts.ca, kensaku@is.nagoya-u.ac.jp

Abstract. Endoscopic navigation generally integrates different modalities of sensory information in order to continuously locate an endoscope relative to suspicious tissues in the body during interventions. Current electromagnetic tracking techniques for endoscopic navigation have limited accuracy due to tissue deformation and magnetic field distortion. To avoid these limitations and improve the endoscopic localization accuracy, this paper proposes a new endoscopic navigation framework that uses an optical mouse sensor to measure the endoscope movements along its viewing direction. We then enhance the differential evolution algorithm by modifying its mutation operation. Based on the enhanced differential evolution method, these movement measurements and image structural patches in endoscopic videos are fused to accurately determine the endoscope position. An evaluation on a dynamic phantom demonstrated that our method provides a more accurate navigation framework. Compared to state-of-the-art methods, it improved the navigation accuracy from 2.4 to 1.6 mm and reduced the processing time from 2.8 to 0.9 seconds.

1 Endoscopic Navigation

Minimally invasive surgery (MIS) is an increasingly performed procedure. Compared to conventional open surgery, it has several benefits for patients, including tiny incisions, faster recovery, and less discomfort. Interventional endoscopy is a minimally invasive procedure commonly used for cancer diagnosis and treatment. The inherent challenge for endoscopists is to exactly locate the endoscope relative to target or suspicious regions in accordance with different image modalities. Pre-operative examination such as computed tomography (CT) scanning is usually performed to analyze anatomical structures of the body before endoscopic interventions. Live endoscopic procedures provide video sequences to visualize the interior of anatomical cavities of the body. Unfortunately, both CT and video images in clinical applications are displayed separately from each other. It is the disconnection of CT and video images that challenges the precise or exact location of the endoscope in clinical practice.

P. Golland et al. (Eds.): MICCAI 2014, Part II, LNCS 8674, pp. 340–348, 2014.

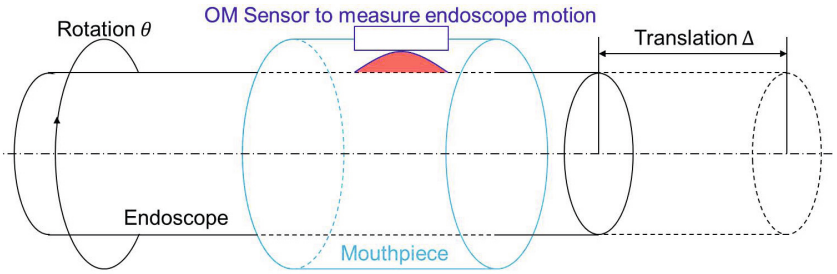


Fig. 1. The external optical mouse tracker that uses the OM sensor to measure rotation θ and translation Δ in the viewing direction of the endoscope motion.

To precisely locate the endoscope navigation methodologies have been developed by fusing preoperative CT slices and real-time video images providing visual guidance in endoscopic interventions. The endoscope localization is generally performed by (1) video-based tracking or (2) electromagnetic localizers.

The video-based navigation performs 2-D/3-D image registration to find the endoscope position [1,2]. It easily fails the navigation if video images are low quality. The EM techniques use an EM position sensor attached at the endoscope tip to measure the endoscope movements. After establishing the alignment between the EM and CT coordinate systems, they can track the endoscope position [3,4]. However, the accuracy of the navigation is limited due to tissue deformations and magnetic field distortions of the EM system.

Beyond current available navigation methods mentioned above, a promising navigation prototype using an optical-mouse (OM) sensor has been discussed recently [5,6]. However, it remains challenging to effectively integrate the OM sensor measurements into endoscopic navigation. This work aims to explore more accurate and robust endoscopic navigation in terms of the OM sensor measurements to determine the endoscope position. The technical highlights of this work are clarified as follows. We proposed an enhanced differential evolution (EDE) algorithm to combine the OM sensor measurements with image structural patches on endoscopic video images to navigate the endoscope. Our proposed framework, which uses the OM sensor, enhanced differential evolution, and endoscopic video images, indicates a new hybrid endoscopic navigation strategy. It not only improves the navigation accuracy but also reduces the processing time. Additionally, it is applicable to different endoscopes, e.g., colonoscope or gastroscope.

2 Approaches

Our navigation framework involves three steps: (1) OM-based motion measurements, (2) EDE for using these measurements, and (3) evaluation of the EDE's population by image structural patches to find the endoscope position.

2.1 Optical Mouse Tracker

The endoscope commonly is manipulated mainly with translation and rotation in its viewing direction. To measure translation and rotation, we use an OM sensor attached at the mouthpiece that is usually utilized during endoscopy (Fig.1). The OM sensor measures motion with 2-D displacements and outputs these displacements, which corresponds to the endoscope translation and rotation.

The OM sensor measurements with their timestamps were collected simultaneously with endoscopic video images. Based on these timestamps, we synchronize the OM sensor measurements and endoscopic video frames. After the synchronization, let $(O_k^{'}, O_k^{''})$ be the OM sensor 2-D outputs at frame k. Relative translation Δ_k and rotational angle θ_k between frames $k-1$ and k are computed:

$$\Delta_k = O_k^{'} - O_{k-1}^{'}, \quad \theta_k = \frac{180^{\circ}(O_k^{''} - O_{k-1}^{''})}{\pi r}, \tag{1}$$

where r is the radius of the endoscope. For more details about the optical mouse tracker to measure the endoscope movements, we refer the reader to the work [6].

2.2 Enhanced Differential Evolution

After obtaining current relative translation Δ_k and rotation θ_k only in the endoscope's viewing direction, we use them to recover the current endoscope's 3-D pose with its position and orientation in six degrees of freedom (6DoF). We propose an enhanced differential evolution algorithm to combine current measured translation Δ_k and rotation θ_k with the current endoscopic video image.

Differential evolution (DE) is a commonly used evolutionary computation method that can find globally optimal solutions to expensive optimization problems through stochastic optimization [7]. It propagates population $\mathcal{P}_{k,g}$ (k indicates the current frame or image) at generation g with N individuals $\mathcal{P}_{k,g} = \{\mathbf{P}_{k,g}^i\}_{i=1}^N$ in G generations (or iterations) for finding the optimal solution. In our endoscopic navigation, individual $\mathbf{P}_{k,g}^i$ is defined as a seven-dimensional (7-D) vector in accordance with the current endoscope 3-D position $\mathbf{t}_k$ and quaternion $\mathbf{q}_k$ which describes the endoscope 3-D orientation:

$$\mathbf{P}_{k,g}^i = (\mathbf{t}_k\ \mathbf{q}_k)_{7\times 1}\,,\ \mathbf{q}_k = (q_{k0}\ q_{k1}\ q_{k2}\ q_{k3})_{4\times 1},\ q_{k0}^2 + q_{k1}^2 + q_{k2}^2 + q_{k3}^2 = 1. \tag{2}$$

During DE, the propagation consists of three operations: (1) mutation, (2) crossover, and (3) selection. The performance of DE depends heavily on the mutation operation as well mutant factor M_i and crossover rate C_i. Unfortunately, the mutation in DE never involves current observation information, which is very important to evolutionary computation. Our strategy to enhance DE is to use currently observed relative translation Δ_k and rotation θ_k for boosting the mutation, as well to employ currently observed video image $\mathbf{I}_k$ to evaluate each individual and adaptively compute mutant factor M_i and crossover rate C_i.

Boosted Mutation. This operation aims to explore the potential solutions during optimization. It generates mutant vector $\mathbf{X}^i_{k,g}$ for each individual $\mathbf{P}^i_{k,g}$ in accordance with mutant factor M_i and difference vectors, $\tilde{\mathbf{X}}^i_{k,g}$ and $\hat{\mathbf{X}}^i_{k,g}$ [7]:

$$\mathbf{X}^i_{k,g} = \mathbf{P}^i_{k,g} + M_i \overbrace{(\mathbf{P}^b_{k,g} - \mathbf{P}^i_{k,g})}^{\tilde{\mathbf{X}}^i_{k,g}} + M_i \overbrace{(\mathbf{P}^{r_1}_{k,g} - \mathbf{P}^{r_2}_{k,g})}^{\hat{\mathbf{X}}^i_{k,g}}, \tag{3}$$

where $\mathbf{P}^b_{k,g}$ is the global best individual at generation g, r_1 and r_2 are mutually exclusive integers randomly selected from $\{1, \cdots, i-1, i+1, \cdots, N\}$.

The mutation operation in Eq. 3 leads to good convergence performance since it uses global best solution $\mathbf{P}^b_{k,g}$. However, global best individual $\mathbf{P}^b_{k,g}$ potentially causes unstable convergence during dynamic optimization since it might deteriorate the population diversity [7]. To address such unstable convergence, we boost the mutation by current observed translation Δ_k and rotation θ_k:

$$\mathbf{X}^i_{k,g} = \mathbf{P}^i_{k,g} + \underbrace{\alpha_i \Omega(\Delta_k, \theta_k)}_{observation} + \tilde{M}_i \tilde{\mathbf{X}}^i_{k,g} + \hat{M}_i \hat{\mathbf{X}}^i_{k,g}, \tag{4}$$

where random number $\alpha_i \in [0, 1]$ controls the preservation of observations Δ_k and θ_k, function $\Omega(\,,\,)$ transforms Δ_k and θ_k to a 7-D vector, and mutant factors $\tilde{M}_i$ and $\hat{M}_i$ will be computed relative to current observed video image $\mathbf{I}_k$. This boosted mutation (Eq. 4) will improve the performance of the DE algorithm.

Crossover. After the mutation, the binomial crossover operation is implemented to determine trial vector $\mathbf{Y}^i_{k,g} = \{y^{i,d}_{k,g}\}_{d=1,\cdots,7}$ in accordance with individual $\mathbf{P}^i_{k,g} = \{p^{i,d}_{k,g}\}_{d=1,\cdots,7}$ and mutant vector $\mathbf{X}^i_{k,g} = \{x^{i,d}_{k,g}\}_{d=1,\cdots,7}$ by

$$y^{i,d}_{k,g} = \begin{cases} x^{i,d}_{k,g} & \text{if } (\mu \le C_i) \text{ or } (d = d_r) \\ p^{i,d}_{k,g} & \text{otherwise} \end{cases}, \tag{5}$$

where μ was randomly generated from an uniform distribution, integer d_r was randomly chosen from $\{1, \cdots, 7\}$, and crossover rate C_i, which checks whether $y^{i,d}_{k,g}$ is copied from $x^{i,d}_{k,g}$, will be calculated relative to the image observation.

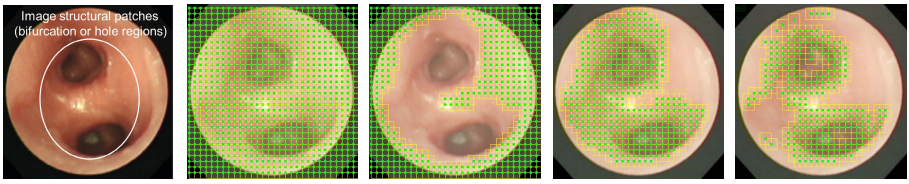
Selection. This operation updates population $\{\mathbf{P}^i_{k,g}\}^N_{i=1}$ to $\{\mathbf{P}^i_{k,g+1}\}^N_{i=1}$ at generation $g+1$ from $\{\mathbf{P}^i_{k,g}\}^N_{i=1} \cup \{\mathbf{Y}^i_{k,g}\}^N_{i=1}$ in terms of their fitness values:

$$\mathbf{P}^i_{k,g+1} = \begin{cases} \mathbf{Y}^i_{k,g} & \text{if } f(\mathbf{Y}^i_{k,g}) \ge f(\mathbf{P}^i_{k,g}) \\ \mathbf{P}^i_{k,g} & \text{otherwise} \end{cases}, \tag{6}$$

where fitness $f(\cdot)$ is also computed on the basis of the current video image.

Eventually, after the G-th iteration, global best solution $\mathbf{P}^*_k$ for the current endoscope 3-D position and orientation at frame k is determined by:

$$\mathbf{P}^*_k = \arg \max_{\mathbf{P}^i_{k,G} \in \{\mathbf{P}^i_{k,G}\}^N_{i=1}} f(\mathbf{P}^i_{k,G}). \tag{7}$$



(a) Input (b) Divided (c) Removed (d) Remaining (e) Selected

Fig. 2. Detect structural patches from input image (a) to selected patches (e): one *yellow* square denotes one patch with a *green* dot indicating its center.

2.3 Image Structural Patches

During EDE discussed above, we must evaluate each individual $\mathbf{P}_{k,g}^i$ in population $\mathcal{P}_{k,g}$, i.e., we must compute fitness $f(\mathbf{P}_{k,g}^i)$ of $\mathbf{P}_{k,g}^i$. We define $f(\mathbf{P}_{k,g}^i)$ as the observation probability relative to current endoscopic video image $\mathbf{I}_k$:

$$f\left(\mathbf{P}_{k,g}^i\right) = \Psi\left(\mathbf{P}_{k,g}^i|\mathbf{I}_k\right). \tag{8}$$

Actually, $\mathbf{P}_{k,g}^i$ is the potential solution to the endoscope 3-D pose parameters. Using $\mathbf{P}_{k,g}^i$ and the CT images, we generate 2-D virtual rendering image $\mathbf{I}_v(\mathbf{P}_{k,g}^i)$ in terms of the surface or volume rendering technique. Hence, it is reasonable to define observation probability $\Psi(\mathbf{P}_{k,g}^i|\mathbf{I}_k)$ as similarity $S(\mathbf{I}_k, \mathbf{I}_v(\mathbf{P}_{k,g}^i))$ between current observed video image $\mathbf{I}_k$ and generated virtual rendering image $\mathbf{I}_v(\mathbf{P}_{k,g}^i)$:

$$\Psi\left(\mathbf{P}_{k,g}^i|\mathbf{I}_k\right) = S\left(\mathbf{I}_k, \mathbf{I}_v(\mathbf{P}_{k,g}^i)\right). \tag{9}$$

The similarity computation should follow the human visual system, which is very sensitive to structural information on images (Fig. 2(a)) [8]. So, we first detect image structural patches (Fig. 2). We divide the image into a set of patches (Fig. 2(b)), use the color information to remove the patches without structures (Fig. 2(c)), and finally select some patches (Fig. 2(e)) from the remaining patches with structures (Fig. 2(d)) to calculate the similarity. Motivated by the work [8], we define similarity $S(\mathbf{I}_k, \mathbf{I}_v(\mathbf{P}_{k,g}^i))$ in the selected image structural patches:

$$S\left(\mathbf{I}_k, \mathbf{I}_v(\mathbf{P}_{k,g}^i)\right) = \frac{1}{2U}\sum_U \left(1 + \frac{(2\xi_k\xi_v + \lambda_1)(2\delta_{k,v} + \lambda_2)}{(\xi_k^2 + \xi_v^2 + \lambda_1)(\delta_k^2 + \delta_v^2 + \lambda_2)}\right), \tag{10}$$

where U is the number of the selected structural patches from the current image. $\delta_{k,v}$ is the correlation between images $\mathbf{I}_k$ and $\mathbf{I}_v(\mathbf{P}_{k,g}^i)$; ξ_k and ξ_v are the average intensity ; δ_k and δ_v are the intensity variances. Two constants λ_1 and λ_2 depends on image intensity and prevents Eq. 10 from unstably calculating the similarity when element $(\xi_k^2 + \xi_v^2)$ or $(\delta_k^2 + \delta_v^2)$ in Eq. 10 is very close to zero.

Based on observation probability $\Psi(\mathbf{P}_{k,g}^i|\mathbf{I}_k)$, we adaptively calculate the mutant factors (Eq. 4) and the crossover rate (Eq. 5) relative to image $\mathbf{I}_k$:

$$\tilde{M}_i = \frac{2\Psi(\mathbf{P}_{k,g}^b|\mathbf{I}_k)}{\Psi(\mathbf{P}_{k,g}^i|\mathbf{I}_k) + \Psi(\mathbf{P}_{k,g}^b|\mathbf{I}_k)},\ \hat{M}_i = \frac{2\Psi(\mathbf{P}_{k,g}^i|\mathbf{I}_k)}{\Psi(\mathbf{P}_{k,g}^i|\mathbf{I}_k) + \Psi(\mathbf{P}_{k,g}^b|\mathbf{I}_k)}, \tag{11}$$

Algorithm 1. Pseudocode of our new endoscopic navigation framework

Input: Endoscopic video images, OM sensor Measurements, the CT images
Output: Endoscope position and orientation $(\mathbf{t}_k^* \ \mathbf{q}_k^*)$ for each video image

for $k = 1$ **to** K (Frame or measurement number) **do**
  ❶ Get current OM measurements – translation Δ_k and rotation θ_k (Eq.1);
  ❷ For current endoscopic video image $\mathbf{I}_k$, detect image structural patches;
  ❸ Compute fitness (Eqs. 8∼10) for individual $\mathbf{P}_{k,g}^i$ and get global best $\mathbf{P}_{k,g}^b$;
  for $g = 1$ **to** G (Generation number) **do**
    for $i = 1$ **to** N (Individual number) **do**
      ❹ Calculate evolutionary parameters, $\tilde{M}_i$, $\hat{M}_i$, C_i (Eqs. 11 and 12);
      ❺ Boosted mutation to determine mutant vector $\mathbf{X}_{k,g}^i$ (Eq. 4);
      ❻ Perform binomial crossover to obtain trial vector $\mathbf{Y}_{k,g}^i$ (Eq. 5);
      ❼ Compute fitness (Eqs. 8∼10), selection (Eq. 6),and update $\mathbf{P}_{k,g}^b$;
    end
    $g = g + 1$;
  end
  ❽ Obtain population $\{\mathbf{P}_{k,G}^i\}_{i=1}^N$ after the G-th generation;
  ❾ Find current best estimate $\mathbf{P}_k^* \longleftrightarrow (\mathbf{t}_k^* \ \mathbf{q}_k^*)$ (Eq. 7);
  $k = k + 1$;
end

$$C_i = \frac{\Psi(\mathbf{P}_{k,g}^i|\mathbf{I}_k) + \Psi(\mathbf{X}_{k,g}^i|\mathbf{I}_k)}{2}. \tag{12}$$

Our endoscopic navigation using the optical mouse tracker, the enhance differential evolution, and structural patches is summarized in **Algorithm 1**.

3 Validation

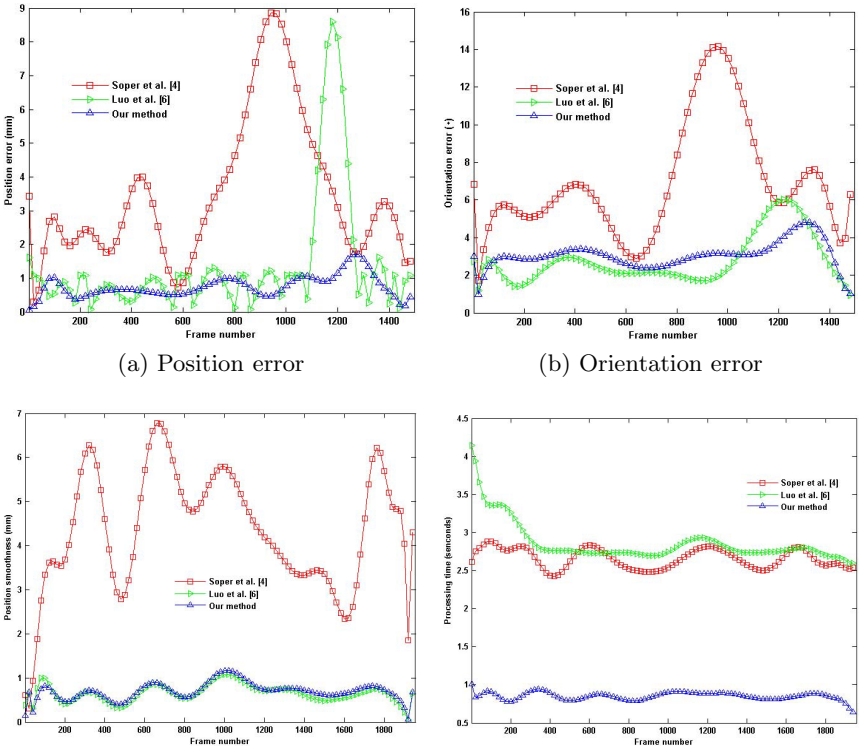
Due to lack of clinical datasets, we evaluate our framework on a dynamic phantom that can simulate tissue deformation. To compare with EM-based navigation, a 3-D Guidance medSAFE tracker (Ascension Technology Corporation, USA) was used as the EM system with the EM sensor (1.5 mm, 6DoF). We also used an endoscope (BF Type P260F, Olympus, Tokyo) to record endoscopic images at video rate 30 frames per second. We investigate th following methods: (1) Soper et al. [4], a hybrid method of combining video- and EM-based tracking, (2) Luo et al. [6]: an enhanced particle filtering method for using OM sensor measurements and endoscopic video sequences, (3) our method, as discussed in Section 2. We set $G = 3$ and $N = 30$. We manually generated ground truth datasets. We calculate the navigation accuracy of different navigation methods.

4 Results and Discussion

Table 1 quantifies the navigation accuracy of using different methods. Our method improved the average accuracy from 2.4 to 1.8 mm. Fig. 3(a)-(b) plot

Table 1. Quantitative comparison of navigation accuracy of different methods

Experiments	Soper et al. [4]	Luo et al. [6]	Our method
1	4.4 mm, 8.6°	2.2 mm, 4.2°	1.6 mm, 4.5°
2	4.5 mm, 8.9°	3.8 mm, 7.2°	3.0 mm, 5.9°
3	3.4 mm, 6.8°	1.1 mm, 2.7°	0.8 mm, 4.2°
4	3.9 mm, 7.2°	1.8 mm, 3.2°	1.1 mm, 3.0°
5	4.2 mm, 8.2°	3.2 mm, 6.2°	2.4 mm, 4.6°
Average	4.1 mm, 7.9°	2.4 mm, 4.7°	1.8 mm, 4.4°



(a) Position error (b) Orientation error

(c) Tracking smoothness (d) Processing time

Fig. 3. (a) and (b) give examples of navigation errors (Experiment 3); (c) tracking position smoothness that was defined as the Euclidean distance between two consecutive frames to evaluate the navigation jump (Experiment 1), (d) the processing time per frame was 2.6, 2.8, and 0.9 seconds by different methods.

the navigation position and errors of the compared methods. Fig. 3(c) shows the tracking smoothness that was improved from 4.8 to 2.0 mm. The processing time per frame was significantly reduced from 2.8 to 0.9 seconds, as shown in Fig. 3(d).

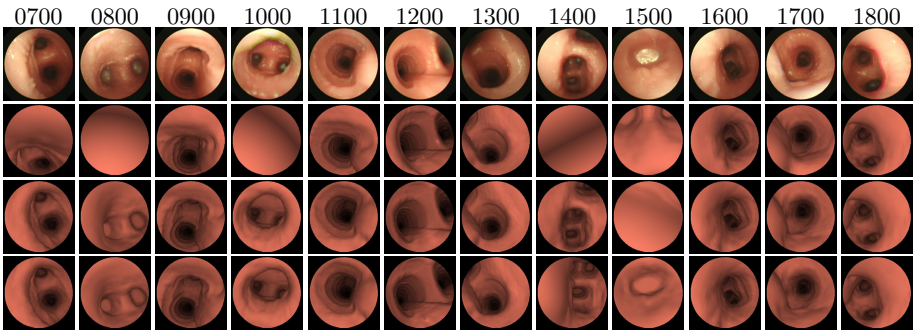


Fig. 4. Visual comparison of the endoscopic video with the virtual images generated during navigation in Experiment 1. The first row of images shows uniformly selected frames from the endoscopic video and the remaining rows show, from top to bottom, the corresponding virtual images generated using Soper et al. [4], Luo et al. [6], and our method. Frame numbers are shown at the top of the figure.

In general, our endoscopic navigation framework outperforms other approaches. Compared to the method [6], we compute the individual's fitness by more accurate structural similarity (Eq.10) that is robust to image luminance and contrast changes. Also, our EDE-based navigation can use less individuals to perform a hierarchical propagation, which not only provides good accuracy but also greatly reduces the computational time of one frame. However, our method may potentially be adversely affected by image artifacts during the fitness computation.

In summary, this work developed a new hybrid endoscopic navigation method using the OM sensor, EDE, and image structural regions. The navigation accuracy was improved while the computational time was also significantly reduced.

Acknowledgment. This work was partly supported by the Japan Society for the Promotion of Science and JSPS Bilateral International Collaboration Grant.

References

1. Luo, X., Feuerstein, M., Deguchi, D., Kitasaka, T., Takabatake, H., Mori, K.: Development and comparison of new hybrid motion tracking for bronchoscopic navigation. MedIA 16(3), 577–596 (2012)
2. Merritt, S.A., Khare, R., Higgins, W.E.: Interactive CT-video registration for the continuous guidance of bronchoscopy. IEEE TMI 32(8), 1376–1396 (2013)
3. 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: Wong, K.H. and Miga, M.I. (eds.) SPIE Medical Imaging 2010, vol. 7625, pp. 76250W, California, USA (2010)
4. Soper, T.D., Haynor, D.R., Glenny, R.W., Seibel, E.J.: In vivo validation of a hybrid tracking system for navigation of an ultrathin bronchoscope within peripheral airways. IEEE TBME 57(3), 736–745 (2010)

5. Cornish, D.C., Higgins, W.E.: Bronchoscopy guidance system based on bronchoscope-motion measurements. In: SPIE Medical Imaging 2012, vol. 8316, pp. 83161G, California, USA (2012)
6. Luo, X., Kitasaka, T., Mori, K.: Externally navigated bronchoscopy using 2-d motion sensors: Dynamic phantom validation. IEEE TMI 32(10), 1745–1764 (2013)
7. Price, K.V., Storn, R.M., Lampinen, J.A.: Differential evolution: A practical approach to global optimization. Springer-Verlag, New York (2005)
8. Wang, Z., Bovik, A.C.: A universal image quality index. IEEE Signal Processing Letters 9(3), 81–84 (2002)
9. Zhang, J., Sanderson, A.C.: JADE: Adaptive differential evolution with optional external archive. IEEE TEVC 13(5), 945–958 (2009)