

# Observation-Driven Adaptive Differential Evolution and Its Application to Accurate and Smooth Bronchoscope Three-Dimensional Motion Tracking

Xiongbiao Luo<sup>a,b,\*</sup>, Ying Wan<sup>c</sup>, Xiangjian He<sup>c</sup>, Kensaku Mori<sup>a</sup>

<sup>a</sup>Information and Communications Headquarters, Nagoya University, Japan;

<sup>b</sup>Robarts Research Institute, Western University, Canada

<sup>c</sup>School of Computing and Communications, University of Technology, Sydney, Australia

---

## Abstract

This paper proposes an observation-driven adaptive differential evolution algorithm that fuses bronchoscopic video sequences, electromagnetic sensor measurements, and computed tomography images for accurate and smooth bronchoscope three-dimensional motion tracking. Currently an electromagnetic tracker with a position sensor fixed at the bronchoscope tip is commonly used to estimate bronchoscope movements. The large tracking error from directly using sensor measurements, which may be deteriorated heavily by patient respiratory motion and the magnetic field distortion of the tracker, limits clinical applications. How to effectively use sensor measurements for precise and stable bronchoscope electromagnetic tracking remains challenging. We here exploit an observation-driven adaptive differential evolution framework to address such a challenge and boost the tracking accuracy and smoothness. In our framework, two advantageous points are distinguished from other adaptive differential evolution methods: (1) the current observation including sensor measurements and bronchoscopic video images is used in the mutation equation and the fitness computation, respectively, and (2) the mutation factor and the crossover rate are determined adaptively on the basis of the current image observation. The experimental results demonstrate that our framework provides much more accurate and smooth bronchoscope tracking than the state-of-the-art methods. Our approach reduces the tracking error from 3.96 to 2.89 mm, improves the tracking smoothness from 4.08 to 1.62 mm, and increases the visual quality from 0.707 to 0.741.

**Keywords:** Adaptive differential evolution, Camera 3-D motion tracking, Evolutionary computation, Bronchoscope tracking and navigation, Surgical instrument tracking and navigation.

---

## 1. Introduction

Medical endoscopes have been widely used in minimally invasive surgery (MIS). They have been integrated with cameras at their distal tip and directly inserted into the body through natural orifices (e.g., mouth and nose) to observe the interior of hollow organs, e.g., sinusopes for sinus inspection, colonoscopes for colon/rectum cancer detection, and angioscopes for examining the lumen of blood vessels.

The bronchoscope is an indispensable tool for lung and bronchus cancer diagnosis, staging, and treatment. It is used to perform diagnostic interventions, e.g., transbronchial lung biopsy (TBLB) or transbronchial needle aspiration (TBNA), to puncture suspicious tumors, and to obtain tissue samples for pathological analysis. However, it is difficult to perform TBLB or TBNA successfully to obtain proper tissue since a bronchoscope provides only two-dimensional (2-D) video images without depth information. Hence, physicians have trouble locating it inside the desired bronchus and easily fail to locate its tip in complex bronchial structures with thousands of bronchi and bifurcations.

Bronchoscope tracking was developed to solve the problem of bronchoscope tip location that physicians are confronted

with. It determines the bronchoscope position and orientation in reference three-dimensional (3-D) coordinate systems, such as computed tomography (CT) and magnetic resonance (MR) image coordinate systems. Various methods to track the bronchoscope have been published in the literature. They are divided into two major categories (or a combination of both): (1) bronchoscope image-registration tracking (BIRT) and (2) bronchoscope electromagnetic tracking (BEMT).

BIRT usually defines an optimization function to minimize the pixel differences between bronchoscopic video images and two-dimensional (2-D) virtual images generated from pre-operative 3-D CT images by volume or surface rendering techniques (Levoy, 1990; Tiede et al., 1990). Helferty et al. reported an image registration method on the basis of normalized mutual information (NMI) for bronchoscope tracking (Helferty et al., 2007). Chung et al. registered video and virtual images to estimate the bronchoscope motion by recovering a bidirectional reflectance distribution function (BRDF) (Chung et al., 2006). Deguchi et al. used a modified mean square error (MSE) measure as the optimization function to spatially synchronize video and virtual images and track the bronchoscope tip (Deguchi et al., 2009), which method was further improved by combining the scale invariant feature transform method (Luo et al., 2012a). Although BIRT works well, it relies heavily on the video image quality that might be degenerated by image artifacts (e.g., mo-

---

\*Corresponding author, Tel: +1-519-697-6081

Email address: xluo@robarts.ca (Xiongbiao Luo)

tion blurring or bronchial bubbles) and also easily gets trapped in the local minima in optimization without a proper initial estimate.

BEMT, which is a very active topic of research as well as the topic of our paper, uses an electromagnetic tracker (EMT) with a position sensor fixed at the bronchoscope tip to directly measure bronchoscope motion (Schwarz et al., 2006). Unfortunately, BEMT accuracy is still unavoidably deteriorated by two drawbacks: (1) sensitivity to location problems due to patient airway deformation (e.g., respiratory motion) and (2) sensor measurement incorrectness due to magnetic field distortion from ferrous metals or conductive material within or close to the working volume. To address these drawbacks, Mori et al. proposed a hybrid method that combines image registration and EMT to track the bronchoscope (Mori et al., 2005), which method was further improved by using Kalman filtering and the condensation algorithm (Soper et al., 2010; Luo et al., 2010b, 2012b). Gergel et al. used particle filtering to compensate for breathing motion (Isard and Blake, 1998; Gergel et al., 2010). Deligianni et al. used a position sensor and a  $pq$ -based registration technique to improve EMT accuracy and stability by modeling respiratory motion with an active shape model (Deligianni et al., 2006). Magnetic field distortion is very difficult to correct without optical tracking (Feuerstein et al., 2009). Even though BEMT has been commercialized for clinical applications (Schwarz et al., 2006), the above drawbacks remain challenging.

This work tackles two disadvantages of BEMT to obtain more accurate and smooth tracking to effectively fuse CT images, bronchoscopic video sequences, and EMT sensor measurements. Generally speaking, BEMT is a multi-modal and uncertain tracking procedure since it uses inaccurate or distorted measurement information, i.e., CT images without recording breathing motion information, bronchoscopic image artifacts, and incorrect or jitter EMT sensor outputs, to track the bronchoscope. Therefore, we require a tracking algorithm that can adapt itself to such inaccuracy and distortion. Currently, as one powerful evolutionary algorithm, differential evolution (DE), which was originally developed by Storn and Price (Storn and Price, 1997), has been increasingly applied as a successful optimization technique to address multi-dimensional complex problems. However, its performance depends heavily on the evolutionary parameters of the mutation factor and the crossover rate. We modify the DE algorithm and propose an observation-driven adaptive differential evolution (OADE) method, which not only determines the evolutionary parameters adaptively on the basis of the current camera image observation information but also combines the current EMT sensor observation information to mutate each individual in a population during optimization. Our OADE strategy very effectively improves the tracking performance of BEMT.

The preliminary work of our OADE approach was presented at the Asian Conference on Computer Vision (ACCV) (Luo and Mori, 2012). This paper is an extension of that ACCV work. Major modifications include more details of our method, the fitness computation, reduction of the computational time by changing the terminating condition of OADE, and more thor-

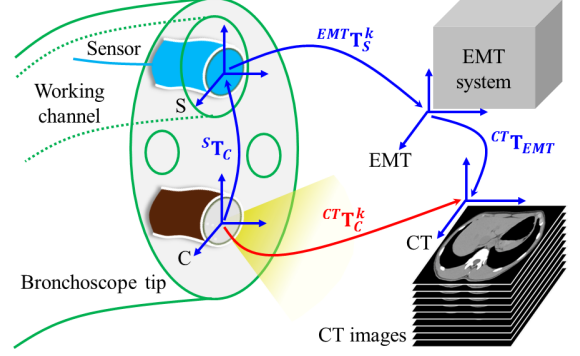


Figure 1: Coordinate systems  $C$ ,  $S$ ,  $EMT$ , and  $CT$  denote bronchoscopic camera, EMT sensor, EMT system, and CT space, respectively. Transformations  ${}^S T_C$ ,  ${}^{EMT} T_S^k$ ,  ${}^{CT} T_{EMT}$ , and  ${}^{CT} T_C^k$  show relationship among them.

ough experiments based on a dynamic phantom. The main contribution of this work is summarized as follows:

- We proposed a new mutation operation for DE methods by integrating the current observation of sensor measurements and camera images, which can control the perturbation velocity and the direction of each individual during evolution, to enhance the DE performance.
- We modified a structural similarity measure to robustly compute the individual's fitness. Since bronchoscopic video images are usually observed with such structural information as bifurcation and fold shapes, we extract these shapes to boost the fitness computation and accurately characterize the individual performance in OADE.
- To the best of our knowledge, our OADE framework is a novel application of DE in bronchoscope 3-D motion tracking. We successfully formulated bronchoscope 3-D motion tracking as an OADE-based stochastic optimization process. EMT sensor measurements, bronchoscopic video images, and CT-based 2-D virtual images can be effectively combined into OADE to achieve a more robust, accurate, and smooth bronchoscope tracking method.
- Additionally, our OADE algorithm is suitable to track other endoscopes, e.g., sinusscopes and colonscopes.

The remainder of this paper is organized as follows. We next formulate the bronchoscope tracking problem in Section 2 and briefly review the basic concepts of the standard different evolution algorithm in Section 3. Our proposed framework is described in Section 4. Section 5 details our experimental settings. Section 6 shows evaluation results and discusses them in Section 7. Finally, our conclusion and future work are shown in Section 8.

## 2. Problem Formulation

This section formulates our tracking problem by parameterizing the bronchoscopic camera motion and generating a 3-D bronchial tree model and 2-D virtual images using CT images.

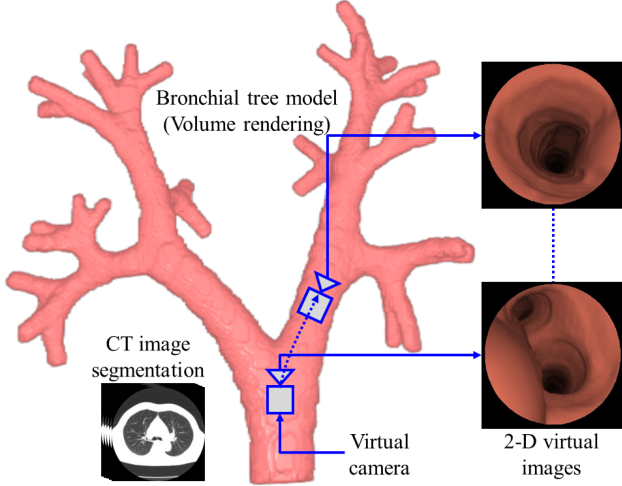


Figure 2: CT image segmentation for obtaining a 3-D bronchial tree model where a virtual camera flies through and generates 2-D virtual images by different camera poses.

### 2.1. Camera 3-D Motion Representation

Bronchoscope 3-D motion tracking determines the bronchoscope tip location including six degrees of freedom (6DoF) position and orientation parameters in CT coordinate systems. It can use different information of bronchoscopic video images, EMT sensor measurements, and CT images. Therefore, such tracking is generally a multi-modal information fusion procedure and involves different coordinate systems (Fig. 1). To integrate the multi-modal information for BEMT, the relationship of these coordinate systems is established by:

$${}^{CT}\mathbf{T}_C^k = {}^{CT}\mathbf{T}_{EMT} {}^{EMT}\mathbf{T}_S^k {}^S\mathbf{T}_C, \quad (1)$$

where  $k$  indicates the  $k$ -th EMT sensor measurement or the  $k$ -th bronchoscopic video image and  ${}^{CT}\mathbf{T}_{EMT}$ ,  ${}^{EMT}\mathbf{T}_S^k$ , and  ${}^S\mathbf{T}_C$  represent different transformation relationships among four coordinate systems of a bronchoscopic camera, an EMT sensor, an EMT system, and CT images during bronchoscope 3-D motion estimation.  ${}^{EMT}\mathbf{T}_S^k$  is the EMT sensor measurement of sensor position  ${}^{EMT}\mathbf{t}_S^k$  and orientation or rotation  ${}^{EMT}\mathbf{R}_S^k$  in the EMT coordinate system.  ${}^{EMT}\mathbf{T}_S^k$  can be represented by a homogeneous matrix:

$${}^{EMT}\mathbf{T}_S^k = \begin{pmatrix} {}^{EMT}\mathbf{R}_S^k & {}^{EMT}\mathbf{t}_S^k \\ \mathbf{0} & \mathbf{1} \end{pmatrix}_{4 \times 4}, \quad (2)$$

which is an inaccurate and distorted measurement of the bronchoscopic camera movement due to the EMT drawbacks.

Bronchoscope motion is characterized by transformation  ${}^{CT}\mathbf{T}_C^k$  that consists of translation or position vector  ${}^{CT}\mathbf{t}_C^k$  and rotation matrix  ${}^{CT}\mathbf{R}_C^k$ , and  ${}^{CT}\mathbf{T}_C^k$  is represented by:

$${}^{CT}\mathbf{T}_C^k = \begin{pmatrix} {}^{CT}\mathbf{R}_C^k & {}^{CT}\mathbf{t}_C^k \\ \mathbf{0} & \mathbf{1} \end{pmatrix}_{4 \times 4}, \quad (3)$$

where  $T$  indicates the transpose operator,  ${}^{CT}\mathbf{t}_C^k = [{}^{CT}t_C^x, {}^{CT}t_C^y, {}^{CT}t_C^z]^T$ , and  ${}^{CT}t_C^x$ ,  ${}^{CT}t_C^y$ , and  ${}^{CT}t_C^z$  are the coordinate values of the bronchoscopic camera position in the  $x$ -,  $y$ -, and  $z$ -axes of the CT coordinate system.

For the rotation part, we used the quaternion but not rotation matrix  ${}^{CT}\mathbf{R}_C^k$  to describe it in our implementation, since the quaternion has been demonstrated to very powerfully characterize the rotation part in accordance with such advantages as the compactness and the avoidance of discontinuous jumps, compared to other representations (e.g., Euler angles). Quaternion  ${}^{CT}\mathbf{Q}_C^k$  to represent rotation  ${}^{CT}\mathbf{R}_C^k$  is a four-element vector (Dam et al., 1998):

$${}^{CT}\mathbf{R}_C^k \longleftrightarrow {}^{CT}\mathbf{Q}_C^k = [Q_0, Q_x, Q_y, Q_z]^T, \quad (4)$$

where the four elements satisfy:  $Q_0^2 + Q_x^2 + Q_y^2 + Q_z^2 = 1$ .

Therefore, bronchoscopic camera motion can be formulated as seven-dimensional (7-D) vector  $\mathbf{X}^k$  at frame  $k$  with respect to position vector  ${}^{CT}\mathbf{t}_C^k$  and quaternion  ${}^{CT}\mathbf{Q}_C^k$ :

$${}^{CT}\mathbf{T}_C^k \longleftrightarrow \mathbf{X}^k = [{}^{CT}\mathbf{t}_C^k; {}^{CT}\mathbf{Q}_C^k]_{7 \times 1}. \quad (5)$$

### 2.2. 3-D Virtual Bronchial Tree Model

During BEMT, the CT image space, which is considered a reference coordinate system, provides a global map where the bronchoscope tip is located. We locate the bronchoscope tip from such a map and segment the CT images to obtain a 3-D virtual bronchial tree model. We imagine a virtual camera flying inside the bronchial tree model. The virtual camera with different position and orientation parameters generates a series of 2-D virtual images by volume rendering techniques (Levoy, 1990) (Fig. 2). Hence, in this sense, BEMT synchronizes the motion between the real bronchoscopic and virtual cameras. BEMT also spatially aligns the current bronchoscopic video image to the 2-D virtual image. After establishing such spatial alignments among 2-D video and virtual images, the virtual camera pose (including position and orientation) exactly corresponds to the bronchoscopic camera location inside the CT image coordinate system. Note that CT-based 2-D virtual images are used to compute the population fitness in our OADE algorithm.

In general, our problem is how to determine transformation  ${}^{CT}\mathbf{T}_C^k$  from the bronchoscopic camera coordinate system to the CT coordinate system using bronchoscopic video sequences, CT-based 2-D virtual images, and EMT sensor measurements. We propose OADE to address this problem.

## 3. Differential Evolution

Next we briefly recall the DE algorithm and introduce some symbols and terms to conveniently explain our OADE method later. After that, we comment on current DE algorithms and show an idea to improve DEs.

Basically, the DE algorithm propagates a population of individuals or vectors  $\{\mathbf{X}_{i,j} | \mathbf{X}_{i,j} \in \mathbb{R}^N\}_{i=1}^P$  ( $P$  is the population size,  $j$  is the generation index,  $j = 1, \dots, G$ , and  $N$  is the dimension of the vector) toward the global optimum during any stochastic optimization procedures. After initializing population  $\{\mathbf{X}_{i,j} | \mathbf{X}_{i,j} \in \mathbb{R}^N\}_{i=1}^P$ , each target vector or individual  $\mathbf{X}_{i,j}$ , which is considered a potential solution to a multi-dimensional

optimization problem, is evolved by performing three operations: mutation, crossover, and selection.

Note that since we parameterize the bronchoscopic camera motion at frame  $k$  as vector  $\mathbf{X}^k$  (Eq. 5), we replace  $\mathbf{X}_{i,j}$  with  $\mathbf{X}_{i,j}^k$  and obtain population  $\mathcal{H}_j^k = \{\mathbf{X}_{i,j}^k | \mathbf{X}_{i,j}^k \in \mathcal{R}^N\}_{i=1}^P$ .

### 3.1. Mutation

The mutation operation expands the potential solutions by updating target vector  $\mathbf{X}_{i,j}^k$  stochastically on the basis of difference vectors and evolutionary factors. The DE performance depends heavily on such an operation. For target vector  $\mathbf{X}_{i,j}^k$  at generation  $j$  at frame  $k$ , its mutant vector  $\mathbf{V}_{i,j}^k$  can be obtained by the frequently used mutation schemes in DEs (Das and Suganthan, 2011):

$$\mathbf{V}_{i,j}^k = \mathbf{X}_{r_1^1,j}^k + F_i (\mathbf{X}_{r_2^2,j}^k - \mathbf{X}_{r_3^3,j}^k) \quad (6)$$

$$\mathbf{V}_{i,j}^k = \mathbf{X}_{best,j}^k + F_i (\mathbf{X}_{r_1^1,j}^k - \mathbf{X}_{r_2^2,j}^k), \quad (7)$$

$$\mathbf{V}_{i,j}^k = \mathbf{X}_{i,j}^k + F_i (\mathbf{X}_{best,j}^k - \mathbf{X}_{i,j}^k) + F_i (\mathbf{X}_{r_1^1,j}^k - \mathbf{X}_{r_2^2,j}^k), \quad (8)$$

$$\mathbf{V}_{i,j}^k = \mathbf{X}_{best,j}^k + F_i (\mathbf{X}_{r_1^1,j}^k - \mathbf{X}_{r_2^2,j}^k) + F_i (\mathbf{X}_{r_3^3,j}^k - \mathbf{X}_{r_4^4,j}^k), \quad (9)$$

$$\mathbf{V}_{i,j}^k = \mathbf{X}_{r_1^1,j}^k + F_i (\mathbf{X}_{r_2^2,j}^k - \mathbf{X}_{r_3^3,j}^k) + F_i (\mathbf{X}_{r_4^4,j}^k - \mathbf{X}_{r_5^5,j}^k). \quad (10)$$

Eqs. 6~10 correspond to the following five strategies: *DE/rand/1*, *DE/best/1*, *DE/target-to-best/1*, *DE/rand/2*, and *DE/best/2*, respectively; they are consistent with a general name or a convention: *DE/a/b*, where *DE* denotes the standard DE algorithm, *a* indicates the base vector to be perturbed, and *b* is the number of difference vectors. Indexes  $r_1^1, r_2^2, r_3^3, r_4^4$ , and  $r_5^5$  are mutually exclusive integers chosen randomly from set  $\{1, \dots, i-1, i+1, \dots, P\}$ . The part in brackets in Eqs. 6~10, e.g.,  $(\mathbf{X}_{r_1^1,j}^k - \mathbf{X}_{r_2^2,j}^k)$ , represents the difference vector;  $\mathbf{X}_{best,j}^k$  is the best individual at generation  $j$ , and  $F_i$  is the mutation factor.

### 3.2. Crossover

The crossover operation improves the diversity of the population and the ability to avoid being trapped into local minima. It is performed to exploit the potential solution space

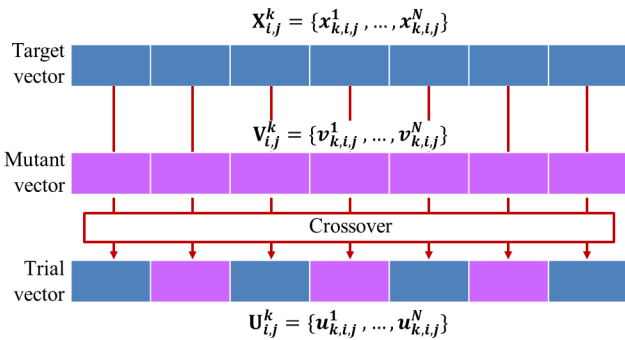


Figure 3: Illustration of crossover step in DE.

### Algorithm 1: Standard DE algorithm

---

```

0. Set population size  $P$ ;
1. Control parameters:  $F_i = \text{constant}$ ,  $C_r = \text{constant}$ ;
2. At generation  $j = 0$ , initialize  $\mathcal{H}_0^k = \{\mathbf{X}_{i,0}^k | \mathbf{X}_{i,0}^k \in \mathcal{R}^N\}_{i=1}^P$ ;
3. Perform the main body of the DE algorithm:
while (termination is unsatisfied)
  for  $i = 1$  to  $P$  do
    ❶ Mutation Operation:
      Generate mutant vector  $\mathbf{V}_{i,j}^k$  for target vector  $\mathbf{X}_{i,j}^k$  using
      one of Eqs. 6~10;
    ❷ Crossover Operation:
      Compute trial vector  $\mathbf{U}_{i,j}^k$  on the basis of vectors  $\mathbf{X}_{i,j}^k$ ,
       $\mathbf{V}_{i,j}^k$ , and Eq. 11;
    ❸ Selection Operation:
      Evaluate  $\mathbf{X}_{i,j}^k$  and  $\mathbf{U}_{i,j}^k$  and choose better vector  $\mathbf{X}_{i,j+1}^k$ 
      by Eq. 12;
  end
   $j = j + 1$ ;
end while

```

---

by exchanging the  $N$ -dimensional information between the target and mutant vectors (Fig. 3). Current DE algorithms use either the exponential or binomial crossover (Das and Suganthan, 2011). In this work, we use the binomial crossover and perform it to generate trial vector  $\mathbf{U}_{i,j}^k = \{u_{k,i,j}^1, \dots, u_{k,i,j}^N\}$  in terms of target vector  $\mathbf{X}_{i,j}^k = \{x_{k,i,j}^1, \dots, x_{k,i,j}^N\}$  and mutant vector  $\mathbf{V}_{i,j}^k = \{v_{k,i,j}^1, \dots, v_{k,i,j}^N\}$ :

$$u_{k,i,j}^d = \begin{cases} v_{k,i,j}^d & \text{if } \text{rand}_d[0, 1] \leq C_r \text{ or } (d = d_{\text{ran}}) \\ x_{k,i,j}^d & \text{otherwise} \end{cases}, \quad (11)$$

where  $\text{rand}_d[0, 1]$  is a random number that yields uniform distribution,  $C_r$  is the crossover rate or the probability that determines whether  $u_{k,i,j}^d \in \mathbf{U}_{i,j}^k$  is copied from  $v_{k,i,j}^d \in \mathbf{V}_{i,j}^k$ , and  $d_{\text{ran}}$  is randomly selected from set  $\{1, 2, \dots, d, \dots, N\}$ .

### 3.3. Selection

The selection step is performed to distinguish whether one target or a trial vector remains or is discarded in the next generation, since the DE algorithm usually maintains the population size during different generations. Such an operation guarantees that all the better individuals are kept in the next generation.

For a maximization problem, the selection operation, which chooses the better individual for the next generation from  $\mathbf{X}_{i,j}^k \cup \mathbf{U}_{i,j}^k$  in terms of their fitness value  $W(\cdot)$ , can be formulated by:

$$\mathbf{X}_{i,j+1}^k = \begin{cases} \mathbf{U}_{i,j}^k & \text{if } W(\mathbf{U}_{i,j}^k, \mathbf{I}^k) \geq W(\mathbf{X}_{i,j}^k, \mathbf{I}^k) \\ \mathbf{X}_{i,j}^k & \text{otherwise} \end{cases}, \quad (12)$$

where  $\mathbf{I}^k$  is the bronchoscopic video image at frame  $k$ .

DE is implemented by the above three operations until the termination is satisfied. The pseudo-code of DE is generalized in **Algorithm 1**. For more details on DE, refer to (Storn and Price, 1997; Das and Suganthan, 2011).



### 3.4. Remarks on DEs

During the last decade, the DE family approaches have been demonstrated to be powerful and easily implemented stochastic optimization algorithms. However, DE performance is controlled by mutation factor  $F_i$  and crossover rate  $C_r$ . Improper factors might collapse the DE's convergence behavior. Small mutation factor  $F_i$  might lead to premature convergence, and large crossover rate  $C_r$  might result in the loss of population diversity. Choosing the best factors to achieve the desired convergence rate is difficult. In particular, different optimization problems may require different factor  $F_i$  and rate  $C_r$  (Das and Suganthan, 2011). Several adaptive differential evolution (ADE) methods were proposed and obtained better performance (Liu and Lampinen, 2005; Brest et al., 2006; Teo, 2006; Qin et al., 2009; Zhang and Sanderson, 2009). In our framework, we also compute these evolutionary parameters adaptively introducing two mutation factors that are based on the individual's fitness value, which is computed by the image intensity differences between the video and the CT-based virtual images, as well as the crossover rate.

On the other hand, to effectively and successfully solve any dynamic or stochastic optimization problems, two general questions must be considered: (1) how to use the current observation information or the temporally/spatially continuous information between two consecutive frames/outputs and (2) how to retain or even enhance the diversity of the population during optimization. In DE, mutation operation answers the first question while the crossover operation deals with the second one. From the point of view of Doucet et al. (Doucet et al., 2000), the optimal solution space of any stochastic optimization problems should be integrated into the current observation. Unfortunately, no current DE or ADE algorithms take the current observation information into account. The mutation operation of DE or ADE disturbs each target vector without any observation information, possibly creating in the local minima or premature convergence problem. To enhance the performance of DEs, we combine the current observation into the mutation operation, as is further discussed in Section 4.

## 4. Observation-Driven Adaptive Differential Evolution

This section explains our OADE framework in greater detail. We formulate the BEMT procedure as a stochastic optimization problem and use our OADE method to solve it.

### 4.1. Method Overview

Our OADE algorithm consists of six steps: (1) preprocessing, (2) initialization and randomization, (3) observation-driven mutation, (4) crossover, (5) individual fitness computation, and (6) determination of the bronchoscope camera motion parameters in 6DoF. The third step is the most advantageous point of our proposed method. We modified the mutation strategy by incorporating the current observation information of the video images and the EMT sensor measurements. We omit discussion of crossover step since it resembles **Algorithm 1**. We also adaptively calculate the mutation factor and the crossover rate

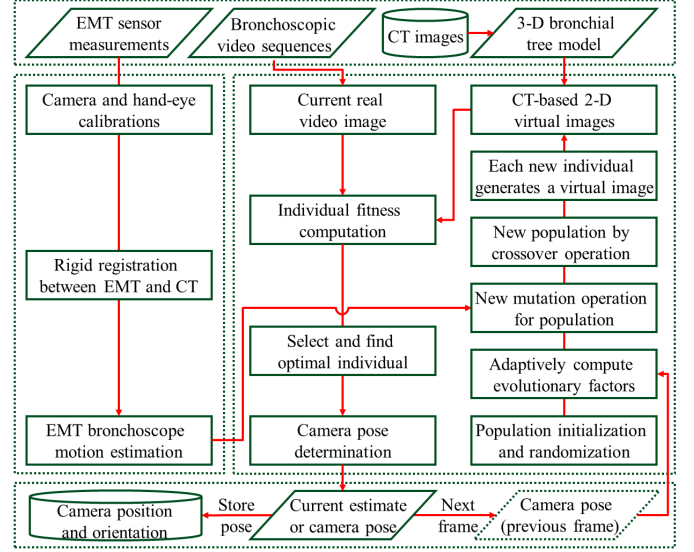


Figure 4: Processing flowchart of our OADE tracking method.

in accordance with the individual's fitness value. In the fifth step, we define a fitness function on the basis of the image intensity difference between the current observation information of the video and CT-based 2-D virtual images. We evaluate each individual in the population and eventually obtain the best individual from the updated population as the current estimate of the camera 3-D motion pose. Fig. 4 illustrates the processing flowchart of our approach with its different steps, which are discussed in the subsequent sections.

### 4.2. Preprocessing

We segment the CT images to obtain the 3-D bronchial tree model. The virtual camera inside it is located differently to generate different 2-D virtual images. In Eq. 1, transformations  ${}^S\mathbf{T}_C$  and  ${}^{CT}\mathbf{T}_{EMT}$  must be determined beforehand using sensor measurement  ${}^{EMT}\mathbf{T}_S^k$ . We perform camera and hand-eye calibrations to compute  ${}^S\mathbf{T}_C$  (Luo et al., 2010a) and use a fiducial-based method to implement a rigid registration to calculate  ${}^{CT}\mathbf{T}_{EMT}$  (Schwarz et al., 2006). Hence, we can obtain bronchoscopic camera motion estimates from EMT. However, these estimates are involved in big errors due to respiratory motion, EMT jitter and magnetic field distortion, and rigid registration for  ${}^{CT}\mathbf{T}_{EMT}$ . Therefore, OADE is performed to improve the BEMT performance and provide more accurate and smooth tracking.

### 4.3. Initialization and Randomization

Our OADE algorithm approximates the global optimal solution in a 7-D space from a randomly initialized population. We use EMT estimate  ${}^{CT}\mathbf{T}_C^1$  (at frame  $k = 1$ ) to initialize each individual in the population in terms of Eq. 5 and obtain  $\mathcal{H}_j^1$  at generation  $j$ :

$$\mathcal{H}_j^1 = \{{}^{CT}\mathbf{T}_C^1 \longleftrightarrow \mathbf{X}_{i,j}^1\}_{i=1}^P. \quad (13)$$

Note that during any stochastic optimization problem, the diversity of the population in the DE methods plays a positive role

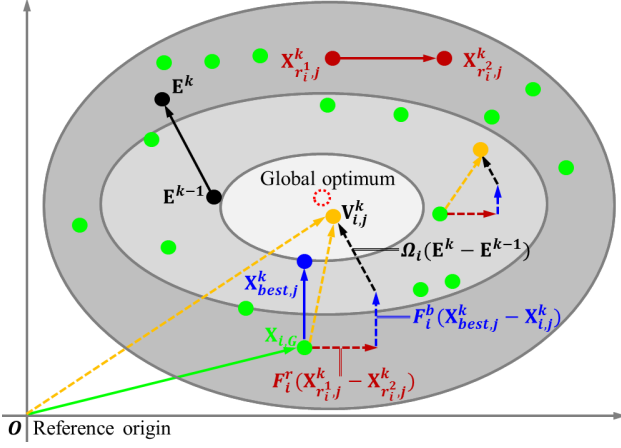


Figure 5: Illustration of our mutation strategy to generate mutant vector  $\mathbf{V}_{i,j}^k$  in accordance with three perturbations including observation  $\Omega_i(\mathbf{E}^k - \mathbf{E}^{k-1})$  and two difference vectors of  $F_i^b(\mathbf{X}_{best,j}^k - \mathbf{X}_{r1,j}^k)$  and  $F_i^r(\mathbf{X}_{r1,j}^k - \mathbf{X}_{r2,j}^k)$ .

in the optimization performance. To enhance the population diversity, we randomize population  $\mathcal{H}_j^1$  in terms of the normal distribution and achieve population  $\mathcal{H}_j^k$  at frame  $k$ :

$$\mathcal{H}_j^k = \{\mathbf{X}_{i,j}^k\}_{i=1}^P, \quad \mathbf{X}_{i,j}^k = \Gamma(\mathbf{X}_{i,j}^1, \pi_i \Delta), \quad (14)$$

where  $\pi_i$  is a normally distributed random number,  $\Delta$  is a pre-defined vector, and  $\Gamma$  is a transform function to add  $\pi_i \Delta$  to  $\mathbf{X}_{i,j}^1$  and obtain  $\mathbf{X}_{i,j}^k$ . Such a randomization is only performed once; i.e., it is only implemented at frame  $k = 1$ .

#### 4.4. New Mutation Operation

The mutation operation is the key step of any DE method. Various research work has focused on how to modify the mutation equation to enhance the standard DE performance (Price et al., 2005; Liu and Lampinen, 2005; Mezura-Montes et al., 2006; Brest et al., 2006; Teo, 2006; Qin et al., 2009; Zhang and Sanderson, 2009). Many mutation strategies were proposed, as shown in Eqs. 6~10. The advantages and limitations of different mutation schemes were discussed (Das and Suganthan, 2011). In our framework, we modify the strategy of  $DE/target - to - best/1$  (Eq. 8) that has good convergence performance due to its usage of the best solution or individual information (Mezura-Montes et al., 2006). However, using the best individual information might result in the loss of population diversity and cause unreliable or precocious convergence. To address such a limitation, we propose a new mutation that combines the current sensor and camera observations to generate mutant vector  $\mathbf{V}_{i,j}^k$  with two difference vectors  $\mathbf{V}_i^b, \mathbf{V}_i^r$  on the basis of  $DE/target - to - best/1$ :

$$\begin{cases} \mathbf{V}_i^b = F_i^b(\mathbf{X}_{best,j}^k - \mathbf{X}_{i,j}^k) \\ \mathbf{V}_i^r = F_i^r(\mathbf{X}_{r1,j}^k - \mathbf{X}_{r2,j}^k) \\ \mathbf{V}_{i,j}^k = \mathbf{X}_{i,j}^k + \underbrace{\Omega_i(\mathbf{E}^k - \mathbf{E}^{k-1})}_{\text{Sensor observation}} + \mathbf{V}_i^b + \mathbf{V}_i^r \end{cases}, \quad (15)$$

where  $\mathbf{E}^k$  and  $\mathbf{E}^{k-1}$  are the EMT estimates on the basis of the EMT sensor measurements at times or frames  $k$  and  $(k-1)$ . Inertia factor  $\Omega_i$  determines how much the current observation is reserved, and we set it to uniformly distributed random number:  $\Omega_i \in [0, 1]$ . On the other hand, we calculate mutation factors  $F_i^b$  and  $F_i^r$  adaptively on the basis of the fitness value of  $\mathbf{X}_{best,j}^k$ ,  $\mathbf{X}_{i,j}^k$  and current camera observation (image)  $\mathbf{I}^k$ :

$$F_i^b = \frac{2W\left(\mathbf{X}_{best,j}^k, \overbrace{\mathbf{I}^k}^{\text{Camera observation}}\right)}{\left(W(\mathbf{X}_{best,j}^k, \mathbf{I}^k) + W(\mathbf{X}_{i,j}^k, \mathbf{I}^k)\right)}, \quad (16)$$

$$F_i^r = \frac{2W(\mathbf{X}_{i,j}^k, \mathbf{I}^k)}{\left(W(\mathbf{X}_{best,j}^k, \mathbf{I}^k) + W(\mathbf{X}_{i,j}^k, \mathbf{I}^k)\right)}. \quad (17)$$

In our new mutation operation, for each individual  $\mathbf{X}_{i,j}^k$ , sensor observation term  $(\mathbf{E}^k - \mathbf{E}^{k-1})$  serves as the external deterministic perturbation that is utilized in terms of random number  $\Omega_i$ . Self-deterministic change  $\mathbf{V}_i^b$  seeks to constrain the distance between the individual and the current best solution. Vector  $\mathbf{V}_i^r$  plays the role of stochastic perturbation. Due to the employment of the current sensor observation, perturbation  $\Omega_i(\mathbf{E}^k - \mathbf{E}^{k-1})$  is very beneficial to lead the population to the best solution space. The self-deterministic and stochastic perturbations are also helpful to diversify the population with various modes. Moreover, we calculate mutation factors  $F_i^b$  and  $F_i^r$  adaptively to control the movements of best difference vector  $(\mathbf{X}_{best,j}^k - \mathbf{X}_{i,j}^k)$  and stochastic difference vector  $(\mathbf{X}_{r1,j}^k - \mathbf{X}_{r2,j}^k)$ . In particular, these factors are related to the individual's fitness that is determined on the basis of the current camera observation information of the video image. In general, our mutation operation introduces the current sensor and camera observations to update the population and computes the two mutation factors adaptively to control the individual's movement, enhancing the exploitation and the exploration abilities of the ADE methods, as will be proved in our experimental results. Fig. 5 shows our mutation procedure to generate vector  $\mathbf{V}_{i,j}^k$ .

During the binomial crossover operation, we automatically update the crossover rate for each individual on the basis of fitness values  $\mathbf{V}_{i,j}^k$  and  $\mathbf{X}_{i,j}^k$ . Since  $C_r$  was suggested within interval  $[0, 1]$  for balancing the global and local searching abilities (Qin et al., 2009), we can calculate it adaptively by:

$$C_r = \frac{\left(W(\mathbf{X}_{i,j}^k, \mathbf{I}^k) + W(\mathbf{V}_{i,j}^k, \mathbf{I}^k)\right)}{2}, \quad (18)$$

which shows a new strategy to control  $C_r$  in terms of fitness  $W(\cdot)$  relative to the current camera image observation.

#### 4.5. Fitness Computation

After the population mutation and crossover, we obtain trial vector set  $\{\mathbf{U}_{i,j}^k\}_{i=1}^P$ . By selecting from sets  $\{\mathbf{X}_{i,j}^k\}_{i=1}^P$  and  $\{\mathbf{U}_{i,j}^k\}_{i=1}^P$ , we can generate target vector  $\mathbf{X}_{i,j+1}^k$  at generation  $(j+1)$ . During the selection operation, the fitness of each target and the

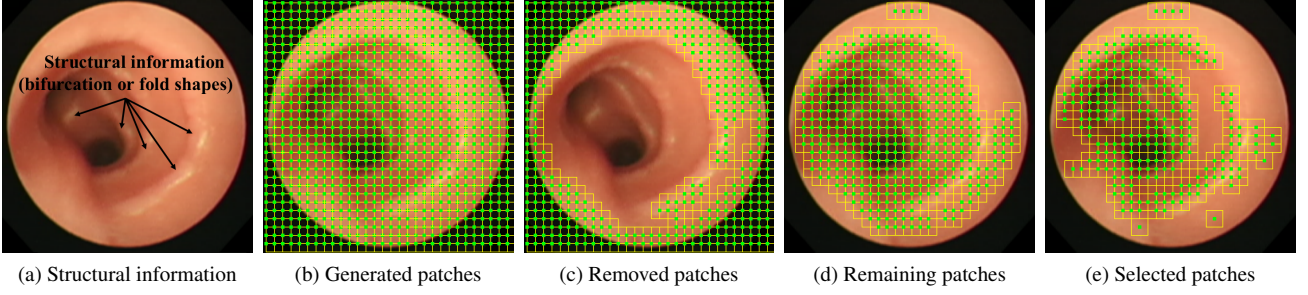


Figure 6: Extract structural regions in one bronchoscopic image to calculate similarity during fitness computation step. One yellow square denotes one patch and a green point is its center.

trial vectors in the population sets must be determined. We define the fitness as the intensity similarity between current video image  $\mathbf{I}^k$  and 2-D virtual image  $\mathbf{I}_v(\mathbf{X}_{i,j}^k)$  (or  $\mathbf{I}_v(\mathbf{U}_{i,j}^k)$ ) generated from the virtual camera with its 6DoF position and orientation parameters of  $\mathbf{X}_{i,j}^k$  (or  $\mathbf{U}_{i,j}^k$ ):

$$W(\mathbf{X}_{i,j}^k, \mathbf{I}^k) = S(\mathbf{I}_v(\mathbf{X}_{i,j}^k), \mathbf{I}^k), \quad (19)$$

where  $S(\cdot)$  is the intensity similarity function. We define  $S(\cdot)$  on the basis of the structural similarity (SSIM) measure (Wang et al., 2004) and explore a modified structural similarity function.

As previously discussed (Wang et al., 2004), separating an image into a number of patches and calculating the similarity locally in each patch is more consistent than directly and globally computing the similarity for the whole image. This is because human observers are usually more sensitive to local statistical features. In bronchoscopic video images, specific structural information, e.g., bifurcations and folds (Fig. 6(a)), is easily perceived by humans. Therefore, extracting the structural information is very useful to compute the similarity. To detect the structural information, we first divide bronchoscopic video image  $\mathbf{I}^k$  (its size  $A \times B$ ) into  $a^* \times b^*$  patches (Fig. 6(b)). All pixel coordinates  $(x, y)$  in patch  $L_{a,b}^k(x, y)$  can be presented by:

$$(\hat{a} - 2)\frac{A}{a^*} \leq x \leq (\hat{a} + 1)\frac{A}{a^*}, (\hat{b} - 2)\frac{B}{b^*} \leq y \leq (\hat{b} + 1)\frac{B}{b^*}, \quad (20)$$

where  $\hat{a}$  and  $\hat{b}$  range with  $2 \leq \hat{a} \leq (a^* - 1)$  and  $2 \leq \hat{b} \leq (b^* - 1)$ . We choose patches from  $\{L_{a,b}^k(x, y)\}$  ( $a = 1, \dots, a^*$ ,  $b = 1, \dots, b^*$ ) in terms of two intensity-based features of patch  $L_{a,b}^k(x, y)$ , standard deviation  $\alpha_{L_{a,b}^k}$  and brightness  $\beta_{L_{a,b}^k}$ :

$$\alpha_{L_{a,b}^k} = \sqrt{\frac{1}{|L_{a,b}^k|} \sum_{(x,y) \in L_{a,b}^k} (\mathbf{I}^k(x, y) - \hat{\mathbf{I}}^k)^2}, \quad (21)$$

$$\beta_{L_{a,b}^k} = \frac{1}{|L_{a,b}^k|} \sum_{(x,y) \in L_{a,b}^k} \Theta_{a,b}(\mathbf{I}^k(x, y)), \quad (22)$$

where  $\hat{\mathbf{I}}^k$  is the image average intensity and  $|L_{a,b}^k|$  is the number of pixels in patch  $L_{a,b}^k$ . Function  $\Theta_{a,b}(\cdot)$  determines the pixel color information of saturation  $\lambda(x, y)$  and lightness  $\sigma(x, y)$  in

the hue-saturation-lightness (HSL) color model.  $\Theta_{a,b}(\cdot)$  determines whether pixel  $(x, y)$  belongs to the highlighted and the lighter areas:

$$\Theta_{a,b}(\mathbf{I}^k(x, y)) = \begin{cases} 1 & \lambda(x, y) \leq \hat{\lambda} \text{ and } \sigma(x, y) \geq \hat{\sigma} \\ 0 & \text{otherwise} \end{cases}, \quad (23)$$

where  $\hat{\lambda}$  and  $\hat{\sigma}$  are two pre-defined thresholds. We remove the white patches without the structural information by  $\beta_{L_{a,b}^k} > \hat{\beta}$  (a fixed constant) (Fig. 6(c)). The remaining patches (Fig. 6(d)) were sorted descendingly in terms of  $\alpha_{L_{a,b}^k}$ . Finally, we select  $\phi = \eta \cdot a^* \cdot b^*$  patches  $\{L_{a,b}^k\}$  (Fig. 6(e)) to calculate the local similarity. Now the similarity between images  $\mathbf{I}^k$  and  $\mathbf{I}_v(\mathbf{X}_{i,j}^k)$  in the selected structural regions  $\{L_{a,b}^k\}$  is computed by:

$$S(\mathbf{I}_v(\mathbf{X}_{i,j}^k), \mathbf{I}^k) = \frac{1}{\phi} \sum_{L_{a,b}^k \in \{L_{a,b}^k\}} \frac{1}{|L_{a,b}^k|} \sum M, \quad (24)$$

where patch similarity measure  $M$  is defined based on the SSIM measure, which was significantly better than a MSE-based measure (Wang et al., 2004). We compute  $M$  by:

$$M = \frac{(2\xi_k\xi_v + C_1)(2\delta_{k,v} + C_2)}{(\xi_k^2 + \xi_v^2 + C_1)(\delta_k^2 + \delta_v^2 + C_2)}, \quad (25)$$

where  $\delta_{k,v}$  is the correlation between  $\mathbf{I}^k$  and  $\mathbf{I}_v(\mathbf{X}_{i,j}^k)$  at patch  $L_{a,b}^k$ ;  $\xi_k$  and  $\xi_v$  are the mean intensity values,  $\delta_k$  and  $\delta_v$  are the intensity variances, and  $C_1$  and  $C_2$  are constants.

In general, Eq. 23 shows a new image similarity measure that very effectively computes the fitness. We detect the structural information from the bronchoscopic video images and calculate the similarity patch by patch to evaluate the individual's performance during our OADE method.

#### 4.6. Camera Pose Determination

Based on the fitness computation and individual selection steps, we obtain new population  $\mathcal{H}_{j+1}^k = \{\mathbf{X}_{i,j+1}^k\}_{i=1}^P$  at generation  $(j + 1)$ . We find optimal solution  $\mathbf{X}_{*,j+1}^k$  with best fitness  $W(\mathbf{X}_{*,j+1}^k, \mathbf{I}^k)$  after each generation and obtain set  $\mathcal{H}^k = \{\mathbf{X}_{*,j+1}^k\}_{j=1}^G$ . From  $\mathcal{H}^k$ , we further choose optimal solution  $\mathbf{X}_{*,*}^k$  in terms of the fitness value for the current bronchoscopic camera motion estimation:

$$\mathbf{X}_{*,*}^k = \arg \max_{\mathbf{X}_{*,j+1}^k \in \mathcal{H}^k} W(\mathbf{X}_{*,j+1}^k, \mathbf{I}^k). \quad (26)$$

---

**Algorithm 2:** OADE-based bronchoscope tracking

---

**Input:** Bronchoscopic video sequences, EMT sensor measurements, and CT images

**Output:** A series of camera 3-D pose (position and orientation in 6DoF)  ${}^{CT}\tilde{\mathbf{T}}_C^k$  in the CT coordinate system

❶ Preprocessing: bronchial model and rigid registration;

❷ Initialization and randomization (Eqs. 13~14);

**for**  $k = 1$  **to**  $K$  (Frame or measurement number) **do**

    Update fitness of each target vector  $\mathbf{X}_{i,j}^k$  using current image  $\mathbf{I}^k$  (Eqs. 19~25), find best target  $\mathbf{X}_{best,j}^k$ ;

**for**  $j = 1$  **to**  $G$  (Generation number) **do**

**for**  $i = 1$  **to**  $P$  (Population number) **do**

            ❸ New mutation: Get  $\mathbf{V}_{i,j}^k$  (Eqs. 15~17);

            ❹ Crossover: Get  $\mathbf{U}_{i,j}^k$  (Eqs. 11 and 18);

            ❺ Fitness computation and selection:

            Compute fitness  $W(\mathbf{U}_{i,j}^k, \mathbf{I}^k)$  of trial vector  $\mathbf{U}_{i,j}^k$  (Eqs. 19~25) and select the better one;

**end**

        Find best target vector  $\mathbf{X}_{*,j+1}^k$  with best fitness  $W(\mathbf{X}_{*,j+1}^k, \mathbf{I}^k)$  from  $\{\mathbf{X}_{i,j+1}^k\}_{i=1}^P$  and store them;

$j = j + 1$ ;

**if**  $W(\mathbf{X}_{*,j+1}^k, \mathbf{I}^k) == W(\mathbf{X}_{*,j+2}^k, \mathbf{I}^k)$  **then**

**break**;

**else**

**continue**;

**end**

**end**

    ❻ Camera pose:  $\mathbf{X}_{*,*}^k \mapsto {}^{CT}\tilde{\mathbf{T}}_C^k$  (Eqs. 26~27);

    Find current best solution  $\mathbf{X}_{*,*}^k$  from  $\mathcal{H}^k = \{\mathbf{X}_{*,j+1}^k\}_{j=1}^G$ ;

$k = k + 1$ ;

**end**

---

Eventually, the output of our OADE method becomes optimal solution  $\mathbf{X}_{*,*}^k = [{}^{CT}\mathbf{t}_C^{k,*} \ {}^{CT}\mathbf{Q}_C^{k,*}]$  relative to bronchoscopic camera motion parameters  ${}^{CT}\tilde{\mathbf{T}}_C^k$  at frame  $k$ :

$$\mathbf{X}_{*,*}^k \mapsto {}^{CT}\tilde{\mathbf{T}}_C^k = \begin{pmatrix} {}^{CT}\tilde{\mathbf{R}}_C^k & {}^{CT}\tilde{\mathbf{t}}_C^k \\ \mathbf{0} & \mathbf{1} \end{pmatrix}. \quad (27)$$

Note that for any optimization problem, a termination condition must be determined. One termination criterion for DE methods is to set generation or iteration number  $G$  as a constant. The optimization automatically stops after  $G$  iterations. On the other hand, the best or optimal fitness is achieved to terminate the execution of a loop. In our case, the fitness (similarity) computation is time-consuming. To reduce the computational time in our OADE processing, we alternately use the termination conditions of the fixed generation number and the optimal fitness. We set the termination condition as  $G$  generations. In OADE, it must render 2-D virtual image  $\mathbf{I}_v(\mathbf{X}_{i,j}^k)$  and compute fitness  $S(\mathbf{I}_v(\mathbf{X}_{i,j+1}^k), \mathbf{I}^k)$  for each individual in population  $\{\mathbf{X}_{i,j+1}^k\}_{i=1}^P$  at generation  $(j + 1)$ . However, virtual image rendering and fitness computation are time-consuming. This

implies that small generation number  $G$  can reduce the computation time. We initially predefine  $G$  as a constant and adjust it automatically during our OADE tracking method. After iterations in the first two generations, we obtain two best fitnesses,  $W(\mathbf{X}_{*,j+1}^k, \mathbf{I}^k)$  and  $W(\mathbf{X}_{*,j+2}^k, \mathbf{I}^k)$ . If  $W(\mathbf{X}_{*,j+1}^k, \mathbf{I}^k) = W(\mathbf{X}_{*,j+2}^k, \mathbf{I}^k)$ , we can achieve the optimal solution after two generations without performing  $(G - 2)$  iterations, since we assume that the remaining iterations can very slightly or do not improve the tracking accuracy any more; otherwise, we execute all  $G$  iterations. Our alternate termination condition is helpful to decrease the runtime under the remaining tracking accuracy of OADE.

Our OADE-based bronchoscope tracking method, which uses EMT sensor measurements, bronchoscopic video images, and CT-based virtual images, is summarized in **Algorithm 2**.

## 5. Experiments

A dynamic phantom that can simulate respiratory motion was constructed to evaluate our proposed method (Luo et al., 2012b). A bronchoscope (BF-P260F, Olympus, Tokyo) was used to acquire  $362 \times 370$  pixel bronchoscopic video images at 30 frames per second. A 3-D Guidance medSAFE Tracker with a 9-coil flat type transmitter (Ascension Technology Corporation, USA) was employed as our EMT system.

We investigate the following EMT-based tracking methods from the literature: (1) Schwarz et al. (Schwarz et al., 2006), directly using absolute EMT sensor outputs; (2) Mori et al. (Mori et al., 2005), integrating absolute EMT sensor outputs into image registration; (3) Luo et al. (Luo et al., 2010a), combining either absolute or relative (inter-frame) EMT sensor outputs with image registration; (4) Luo et al. (Luo et al., 2012b), utilizing sequential Monte Carlo (SMC) methods to fuse inter-frame EMT sensor outputs with video image; (5) Zhang et al. (Zhang and Sanderson, 2009), a method of adaptive differential evolution with optional external archive; (6) our method, as discussed in Section 4. Note that we compared SMC (Luo et al., 2012b) and our proposed to show different optimization frameworks for bronchoscope tracking. We also compared our modified similarity function (Eq. 24) and a modified mean squared error (MoMSE) measure (Deguchi et al., 2009) for fitness computation on the basis of our OADE tracking framework to investigate different similarity measures on bronchoscope motion estimation. Additionally, we point out that it is somewhat hard to thoroughly and fairly compare different differential evolution approaches due to different experimental setups, implementations, and applications. We here compare OADE to a previous method (Zhang and Sanderson, 2009).

We performed 21 experiments (a total of 38,248 bronchoscopic video frames) within six ground truth datasets (GTDs) that were generated in terms of a previous method (Luo et al., 2012b). Additionally, during the fitness computation (Section 4.5), we set parameters  $a^* = b^* = 30$ ,  $\eta = 0.3$ ,  $\hat{\lambda} = 0.6$ ,  $\hat{\delta} = 0.7$ , and  $\hat{\beta} = 0.9$  in accordance with previous work (Deguchi et al., 2009).

We define the tracking error, smoothness, and visual quality to evaluate the results of each method. Based on GTDs, we



compute the position and orientation errors ( $\tau, \theta$ ) by:

$$\tau = \|\mathbf{t} - \mathbf{t}_G\|, \theta = \arccos((\text{trace}(\mathbf{R}\mathbf{R}_G^T) - 1)/2), \quad (28)$$

where  $\tau$  is the Euclidean distance between ground truth position  $\mathbf{t}_G$  and estimate  $\mathbf{t}$  and  $\theta$  is the rotation difference between ground truth rotation  $\mathbf{R}_G$  and estimate  $\mathbf{R}$  by the six tracking methods.

It is also important to evaluate whether the tracking procedure is smooth during the bronchoscopic interventions. Smooth tracking implies that the tracking results include little jitter or jump errors. It can provide two benefits, including reduction of physician discomfort and distraction in operations and enhancement of such posterior procedures as respiratory motion modeling and clinical comparison by temporally and spatially aligning different video sequences from several examinations on patients. We define the smoothness as the average Euclidean distance and the standard deviation of the estimated positions among continuous frames as well as the orientations:

$$\mu = \frac{\sum_{k=1}^{K-1} \|\mathbf{t}_{k+1} - \mathbf{t}_k\|}{K-1}, \quad (29)$$

$$\psi = \frac{\sum_{k=1}^{K-1} \arccos((\text{trace}(\mathbf{R}_{k+1}\mathbf{R}_k^T) - 1)/2)}{K-1}, \quad (30)$$

where  $K$  is the frame number. Large  $\mu$  and  $\psi$  indicate that large jitter or jump occurred in the tracking results.

The tracking results are usually represented by mathematical numbers, i.e., a series of transformation  ${}^{CT}\hat{\mathbf{T}}_C^k$  with the position and orientation parameters in 6DoF. We use these parameters to generate a sequence of 2-D virtual images that can be directly visualized by physicians. We can check whether these virtual images resemble the video images and assess the tracking visualization quality. We define visual quality  $\gamma$  on the basis of a universal image quality index (Wang and Bovik, 2002):

$$\gamma = \frac{1}{2} \left( 1 + \frac{4\vartheta_{k,v}\epsilon_k\epsilon_v}{(\vartheta_k^2 + \vartheta_v^2)(\epsilon_k^2 + \epsilon_v^2)} \right), \quad (31)$$

where  $\vartheta_{k,v}$  is the correlation between image  $\mathbf{I}^k$  and its virtual image  $\mathbf{I}_v$  corresponding to estimated pose  ${}^{CT}\hat{\mathbf{T}}_C^k$ ,  $\vartheta_r$  and  $\vartheta_v$  are the covariance, and  $\epsilon_k$  and  $\epsilon_v$  are the average values.

## 6. Results

In our OADE tracking method, we experimentally determine population size  $P$  and generation number  $G$ . Fig. 7 shows the tracking error and processing time results under different population sizes and generation numbers. We used population size  $P = 25$ . According to Fig. 7(c), we can initially set generation number  $G = 3$  for obtaining the best tracking accuracy. As discussed above (Section 4.6), we automatically adjust  $G$  to reduce the runtime during iterations. The generation number is not adjusted, i.e.,  $G = 3$ , when the best individual fitness of the first generation does not equal the second generation; otherwise,  $G = 2$ . Our choices of  $P = 25$  and  $G = 2$  or 3 can balance the accuracy and runtime.

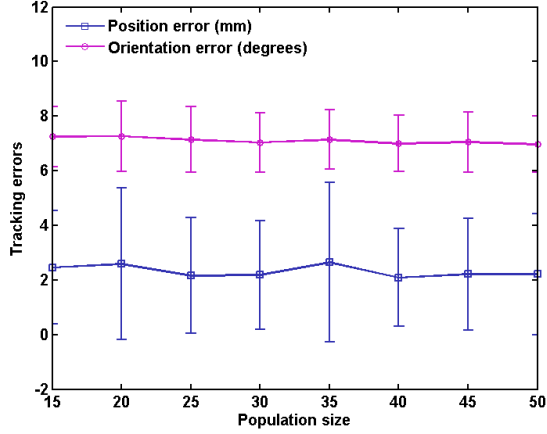
Fig. 8 displays the tracking errors of using the six approaches in Experiments (or Datasets) 5 and 20. Fig. 9 shows an example of the tracking smoothness of Experiment 6. Fig. 10 illustrates the tracking visual quality of using the six methods in Experiments 8 and 14. Table 1 summarizes the average accuracy, the smoothness, and the visual quality of the tracking results estimated by the six methods. The average tracking errors of the position and the orientation of our OADE method were about 2.89 mm and 8.18°, respectively. However, our previous methods provided tracking error of at least 3.96 mm and 10.47°. The tracking smoothness of the position and the orientation were also significantly improved from (4.08 mm, 2.65°) to (1.62 mm, 2.11°). Our proposed OADE framework significantly outperforms other tracking methods. Fig. 11 investigates different optimization frameworks and similarity measures for bronchoscope tracking. The OADE algorithm works better than SMC (Luo et al. (Luo et al., 2012b)) on the basis of the MoMSE measure for fitness computation. Our proposed similarity measure for fitness computation outperforms MoMSE on the basis of our OADE framework.

Fig. 12 compares the tracking smoothness of the six methods during 21 experiments. Fig. 13(a) shows the visual quality of each method in them. Fig. 13(b) investigates the computational times of each method. The processing times of the five methods of Mori et al. (Mori et al., 2005), Luo et al. (Luo et al., 2010a), Luo et al. (Luo et al., 2012b), and Zhang et al. (Zhang and Sanderson, 2009) were about 0.69, 1.13, 1.83, 0.97, and 0.92 seconds per frame (spf), respectively. Compared to (Luo and Mori, 2012), we reduced the average runtime from 1.63 to 0.92 spf, since we reduced the population size from 30 to 25 and alternatively used the generation number ( $G = 2$  or  $G = 3$ ). Note that the method of Schwarz et al. (Schwarz et al., 2006) is real-time processing.

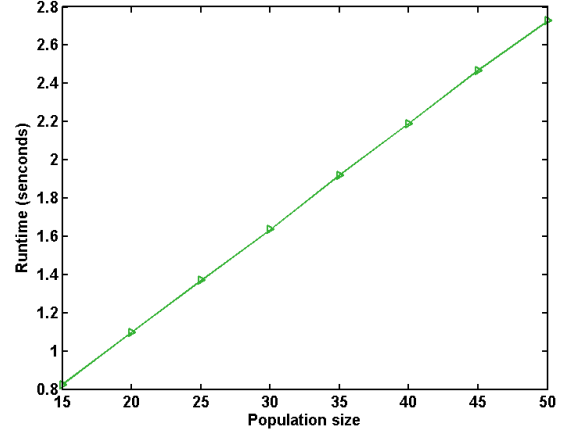
Fig. 14 illustrates the estimated camera trajectories of the methods of Luo et al. (Luo et al., 2012b) and our OADE framework on a pre-built 3-D bronchial tree model. Fig. 15 shows visual comparisons of the real and virtual bronchoscopic images at uniformly selected frames from different approaches. All the 2-D virtual bronchoscopic images were generated from the virtual camera poses (position and orientation) estimated by the six methods. Some virtual images in Fig. 15 were generated with black regions (in circles) since the virtual camera with its estimated pose was outside the bronchial tree volume. Both Figs. 14 and 15 further prove that our OADE method estimates bronchoscope camera movements significantly better than the other methods.

## 7. Discussion

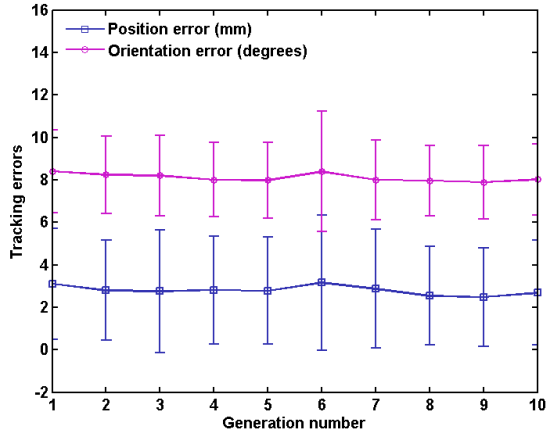
In general, our OADE method provides a more accurate and smooth tracking framework than other available methods to estimate the bronchoscopic camera 3-D motion. The experimental results demonstrate that our method successfully deals with the problems of bronchoscopic image artifacts (e.g., due to specular or inter-reflection, motion blurring, and illumination changes) and EMT sensor measurement inaccuracy due to respiratory motion and the magnetic field distortion of the



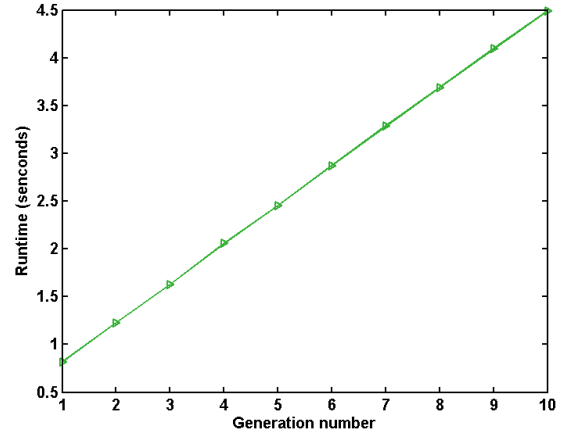
(a) Error of different population sizes



(b) Time of different population sizes



(c) Error of different generation numbers



(d) Time of different generation numbers

Figure 7: Experimentally determined population size and generation number:  $P = 25$  and  $G = 2$  or  $3$ .

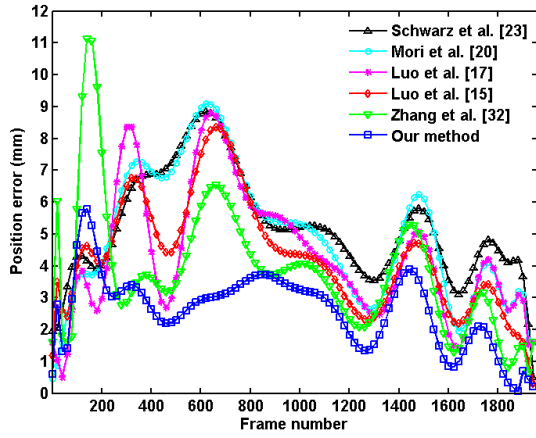
EMT system during bronchoscope tracking. The improvement of tracking accuracy and the smoothness of our OADE method is attributed to several advantages, as discussed as follows.

### 7.1. Effectiveness

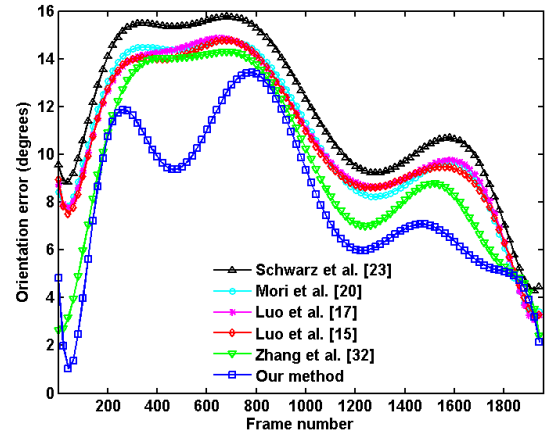
Our OADE method's effectiveness contributes to several advantageous aspects. The main advantage is that our mutation operation introduces current sensor and camera observations (sensor measurements and video images) to update the population, adaptively computes the two mutation factors to control the individual's movement, and automatically calculates the crossover rate. Such mutation enhances the OADE's exploitation and exploration abilities and prevents it from premature convergence. The exploitation ability means that sufficient fitness values occurred in a generation to find even better solutions from the good ones. The larger the individual fitness, the greater is the individual's exploitation. The exploration ability depends on whether all the individuals are distributed sufficiently to cover the solution space and find the global optimum. It also means the population diversity. The more diverse the

individual, the more potential solutions can be obtained. Integrating the current observation information into the mutation operation can propagate the individual correctly to enhance the exploitation and exploration abilities for searching for the best solution. Due to simulated breathing motions and EMT sensor jitter errors, our previous method unsuccessfully addressed the problems, but our OADE can track the bronchoscope successfully and correctly align the video and virtual images. On the other hand, due to the OADE method's robustness in the fitness computation, it can also tackle bronchoscopic image artifacts, as illustrated in Fig. 16.

Furthermore, for the current DE and ADE algorithms, only stochastic perturbation is performed in the mutation step, which may prohibit the population from reaching the current bronchoscope location. However, our modified mutation, which combines the stochastic perturbation with the deterministic perturbation of the current EMT sensor output, can positively guide individuals in the population to approximate the best solution space for the current bronchoscope location. Computing the crossover rate on the basis of the current observation is also

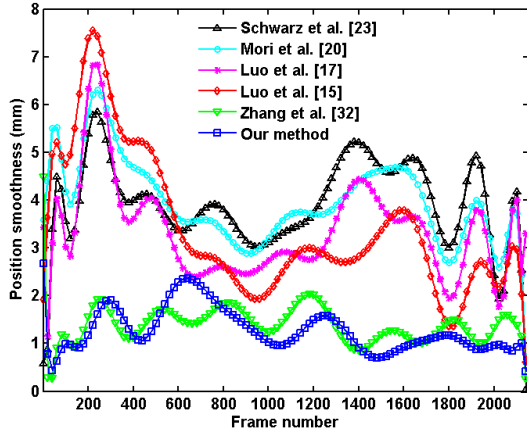


(a) Position error of Experiment 20

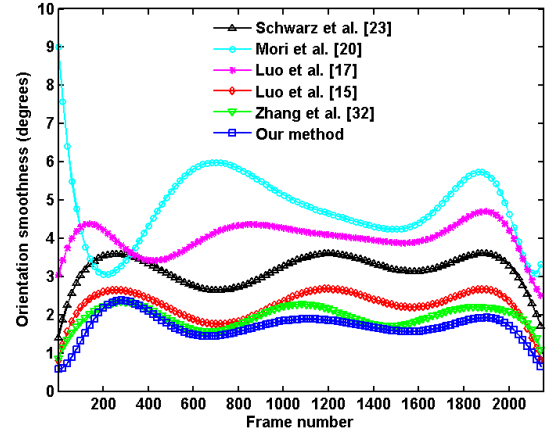


(b) Orientation error of Experiment 20

Figure 8: Plotted tracking error of six methods evaluated in Experiment 20.

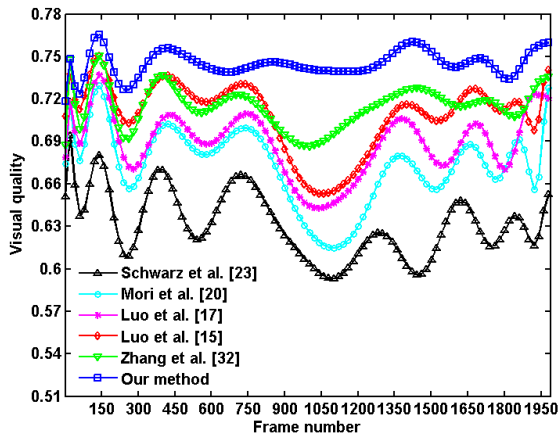


(a) Position smoothness

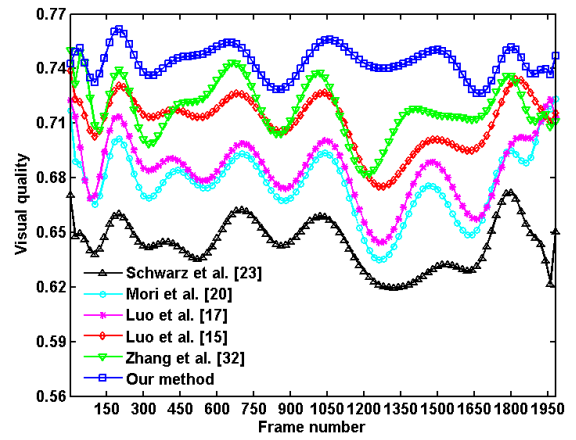


(b) Orientation smoothness

Figure 9: Plotted tracking smoothness of six methods evaluated in Experiment 6.



(a)



(b)

Figure 10: Plotted visual quality of six methods evaluated in Experiments 8 and 14.

Table 1: Comparatively quantified average accuracy, smoothness, and visual quality of tracking results of six methods evaluated in six ground truth datasets. They were calculated in terms of Eqs. 28~31. Our OADE outperformed other methods.

Compared approaches	Average accuracy		Average smoothness		Average visual quality $\gamma$
	Position $\tau$ (mm)	Orientation $\theta$ ( $^\circ$ )	Position $\mu$ (mm)	Orientation $\psi$ ( $^\circ$ )	
Schwarz et al. (2006)	$5.10 \pm 2.65$	$11.46 \pm 8.83$	$4.67 \pm 0.90$	$3.57 \pm 0.39$	$0.636 \pm 0.008$
Mori et al. (2005)	$4.43 \pm 3.29$	$10.98 \pm 9.45$	$4.80 \pm 0.80$	$4.91 \pm 0.35$	$0.668 \pm 0.086$
Luo et al. (2010a)	$4.39 \pm 4.23$	$10.63 \pm 9.94$	$4.27 \pm 0.86$	$4.29 \pm 0.36$	$0.680 \pm 0.094$
Luo et al. (2012b)	$3.96 \pm 3.60$	$10.47 \pm 8.82$	$4.08 \pm 0.59$	$2.65 \pm 0.33$	$0.707 \pm 0.095$
Zhang and Sanderson (2009)	$3.63 \pm 3.21$	$9.53 \pm 7.55$	$1.93 \pm 0.58$	$2.35 \pm 0.96$	$0.715 \pm 0.004$
Our method	$2.89 \pm 2.87$	$8.18 \pm 6.29$	$1.62 \pm 0.53$	$2.11 \pm 0.69$	$0.741 \pm 0.005$

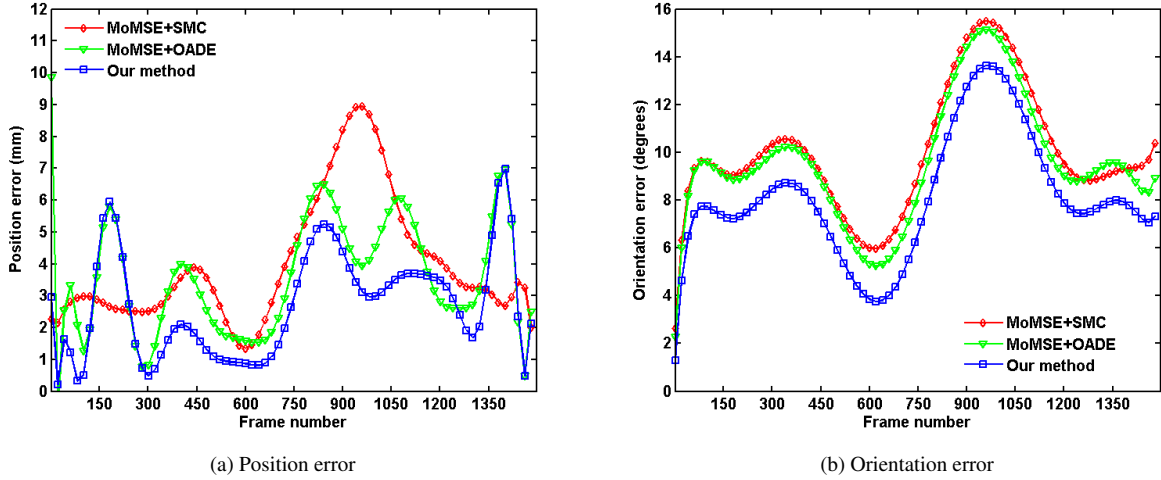


Figure 11: Comparison of different similarity measures and optimization frameworks for 3-D bronchoscope motion estimation in Experiment 9. Tracking error of the MoMSE+SMC method (Luo et al., 2012b) was 3.95 mm and  $9.99^\circ$  while the MoMSE+OADE approach was 3.50 mm and  $9.58^\circ$ . Our method shows best performance with 2.69 mm and  $8.09^\circ$ .

helpful to maintain the population diversity. In contrast to our previous SMC-based tracking method (Luo et al., 2012b), OADE has more powerful exploitation and exploration abilities since it can maintain the multiplicity of population and augment the capacity of local searches, while SMC is somewhat constrained on the diversity loss (Fig. 16).

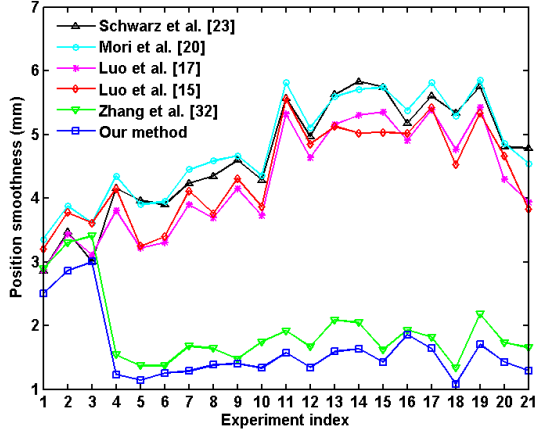
Finally, we believe that our modified similarity measure for fitness computation also enhances the tracking performance. It is important to accurately evaluate individual performances by computing fitness during iterations. Extracting the structural information to calculate the fitness is beneficial to correctly characterize individual performances. Our previous method used a MSE-based measure to evaluate them. The OADE tracking approach works much better than the MSE-based method (Luo et al., 2012b) since the structural similarity measure was proved to significantly outperform MSE (Wang et al., 2004).

## 7.2. Limitations

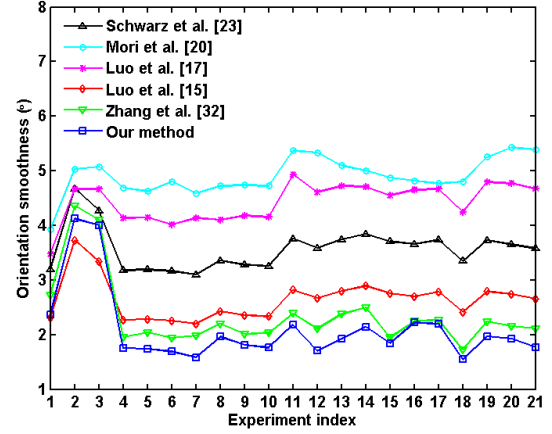
Although our proposed OADE-based tracking approach significantly outperforms the state-of-the-art methods, it still has several potential limitations. The first open issue is its computational time that requires about 0.92 seconds to process one frame, which is slower than the camera video rate of 30 frames

per second. The main time was spent computing the fitness values between the video and virtual images. The fitness computation for each individual in the population involves both generating 2-D virtual images by volume rendering techniques and computing the similarity (Eq. 24). The processing times of both parts take around 0.65 and 0.26 seconds, respectively. 2-D virtual image generation greatly slows down our OADE tracking. The similarity computation also takes much time since it processes image intensity pixel by pixel. The processing time of our method challenges current computer devices. We still exert much effort to improve the computational efficiency in our future work. One option is to replace volume rendering by surface rendering to reduce the processing time. Another way to improve the computational efficiency is to reduce the number of individuals and generations and develop a more robust fitness computation method to characterize individuals after evaluation. It might also be helpful to use a graphics processing unit (GPU) or multi-threading techniques to accelerate processing. We also clarified that this work focuses on improving continuous tracking accuracy to meet the clinical requirement of 2.0 mm during bronchoscopic navigation. We may sacrifice computational efficiency to enhance the performance of continuous bronchoscopic navigation. In some cases, e.g., when the bron-



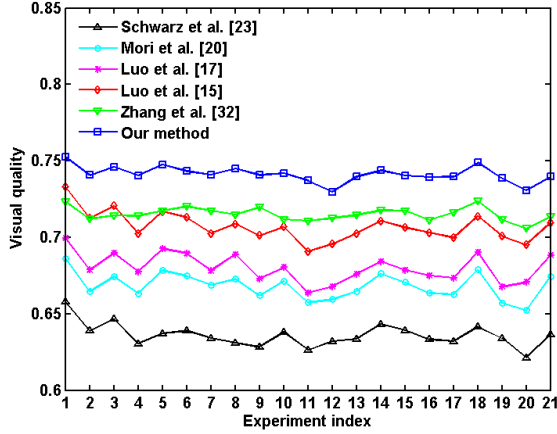


(a) Position smoothness

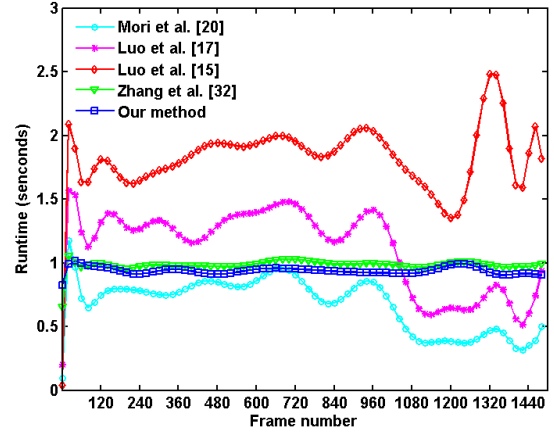


(b) Orientation smoothness

Figure 12: Comparison of tracking (position and orientation) smoothnesses of six methods during 21 experiments.



(a) Visual quality



(b) Computational time

Figure 13: Visual quality of six methods during 21 experiments and computational time of five methods.

choscope has been already navigated to the target regions at the correct airway, we could reduce the computational time by sacrificing tracking accuracy.

We implemented a relatively simple image processing method that uses HSL model-based color information detection and involves several predefined thresholds to identify the structural regions in our fitness computation, possibly resulting in incorrectly calculated fitness. We consider either inter-pixel similarity or more robust functions (to replace Eq. 23) without defining the thresholds to perform patching or aggregation.

Additionally, we did not evaluate our method on clinical datasets. Of course, such validation is optimal. Unfortunately, obtaining them is really difficult since limited clinical datasets exist worldwide, even though a commercially available EMT-based bronchoscopy system, superDimension, has been increasingly introduced in operating rooms. We also point out that the current tracking error is relative error since we defi-

nately introduce error in our ground truth datasets. Obtaining real ground truth data is very challenging. The current tracking error on the basis of our dynamic phantom evaluation might become larger than validating our method on clinical datasets.

In general, our proposed OADE tracking method suffers from its computational time, which currently limits its clinical applications or practice. Clinical or non-phantom validation must also be performed before introducing the OADE tracking framework into operating rooms.

## 8. Conclusion

This work exploited an observation-driven adaptive differential evolution approach to accurately and smoothly track the bronchoscope tip (camera) motion. By integrating the current sensor and camera observation information in the mutation operation and computing the mutation factor and the crossover

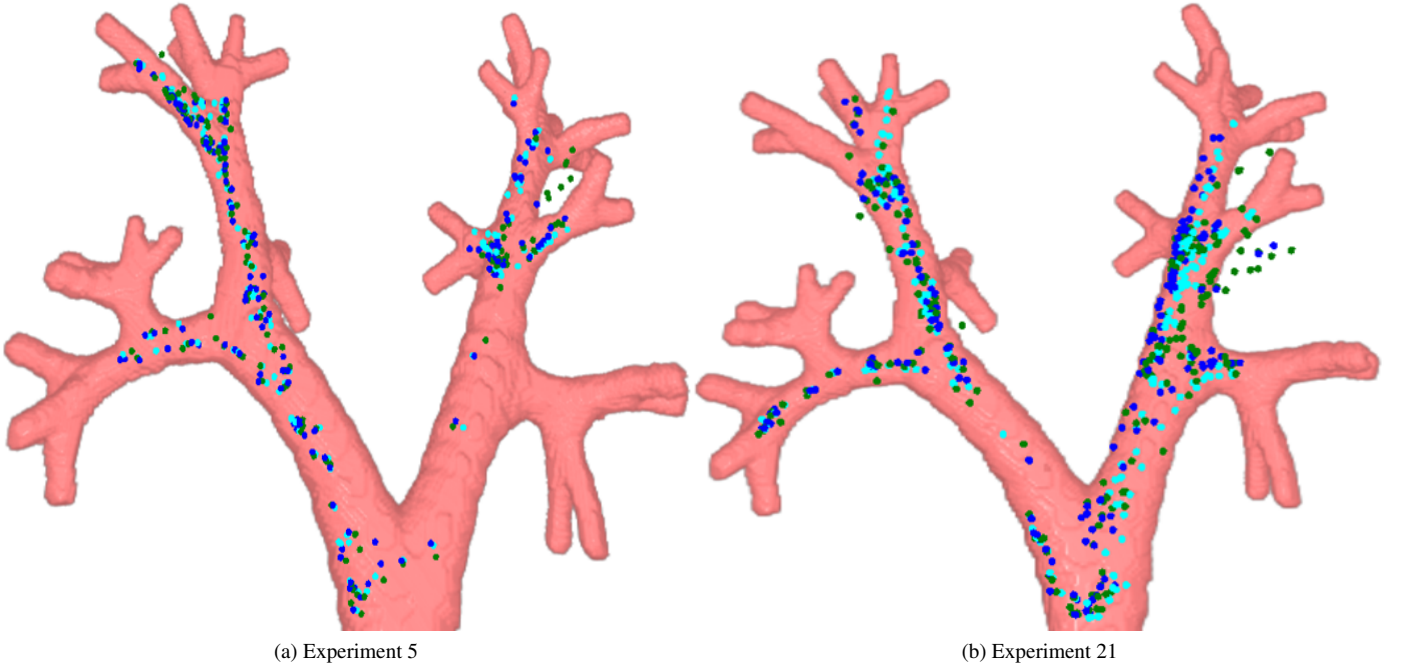


Figure 14: Plotted tracking results (Experiments 5 and 21) as camera motion paths on pre-built 3-D bronchial tree model using our proposed OADE method (*blue dots*) and the method of Luo et al. (Luo et al., 2012b) (*green dots*). Dots demonstrate that our method overlaps more ground truth dots or follow longer camera movement paths than Luo et al. (Luo et al., 2012b) (*cyan dots* show ground truth).

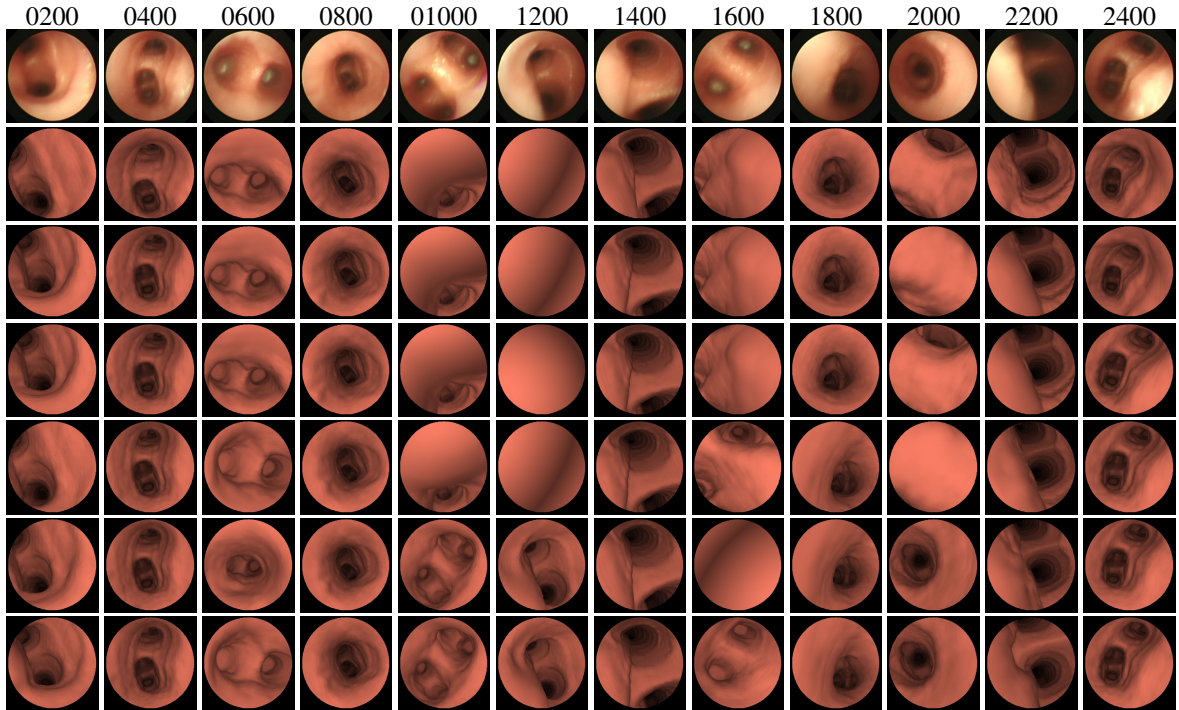
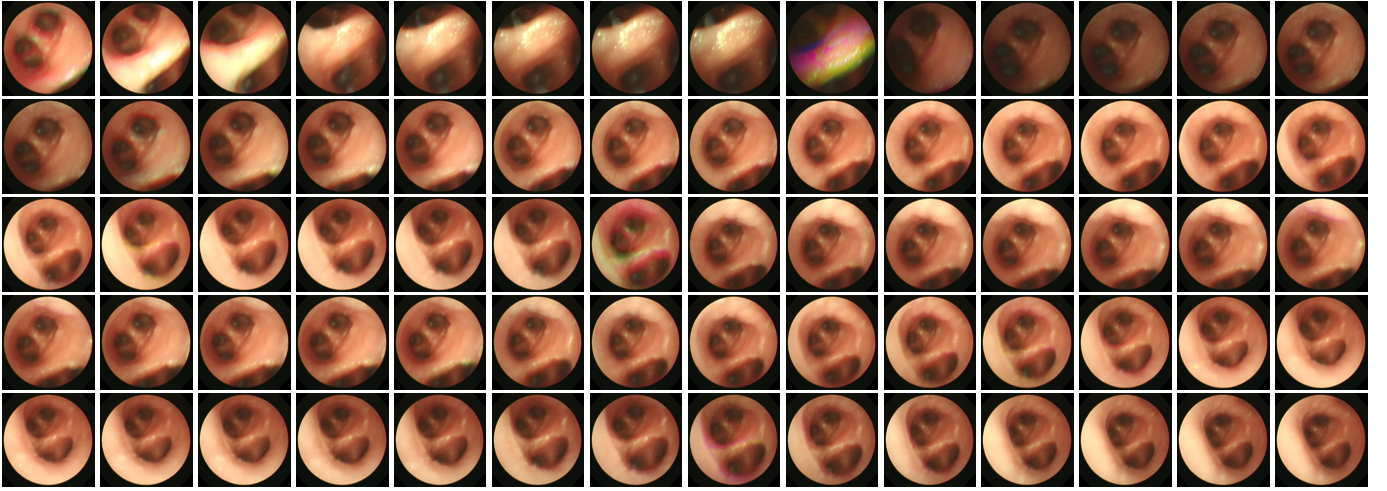
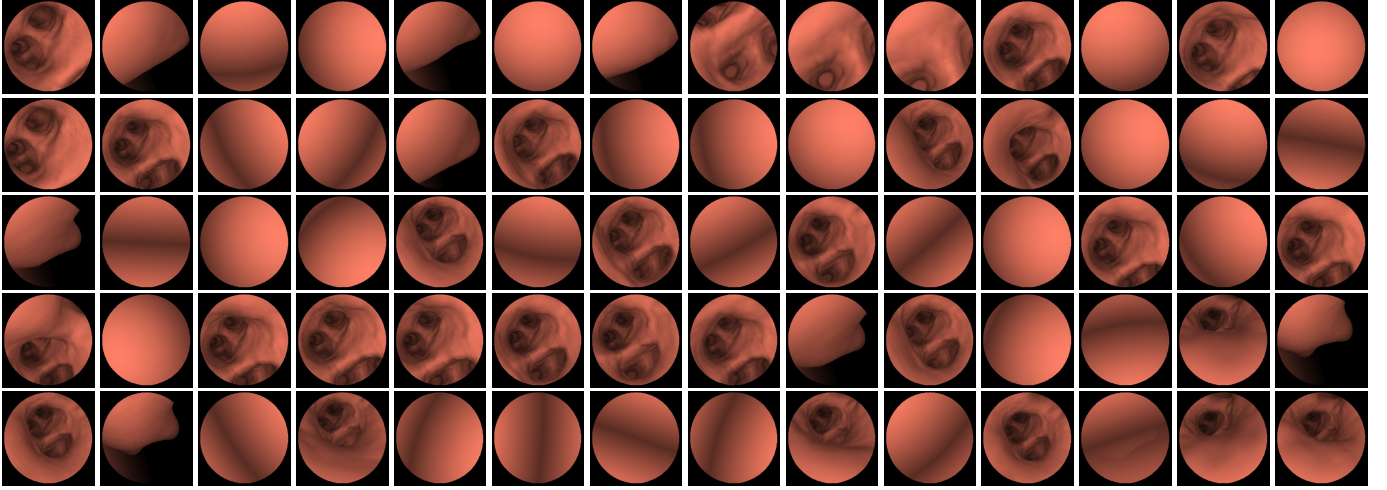


Figure 15: Examples of visual comparison of tracking results of Experiment 18. Top row shows frame numbers selected uniformly every 200 frames, and second row shows corresponding real images. Other rows (from third to sixth) display virtual images generated from tracking results using methods of Schwarz et al. (2006), Mori et al. (2005), Luo et al. (2010a), Luo et al. (2012b), Zhang and Sanderson (2009), and ours, respectively. Our proposed framework shows the best performance.

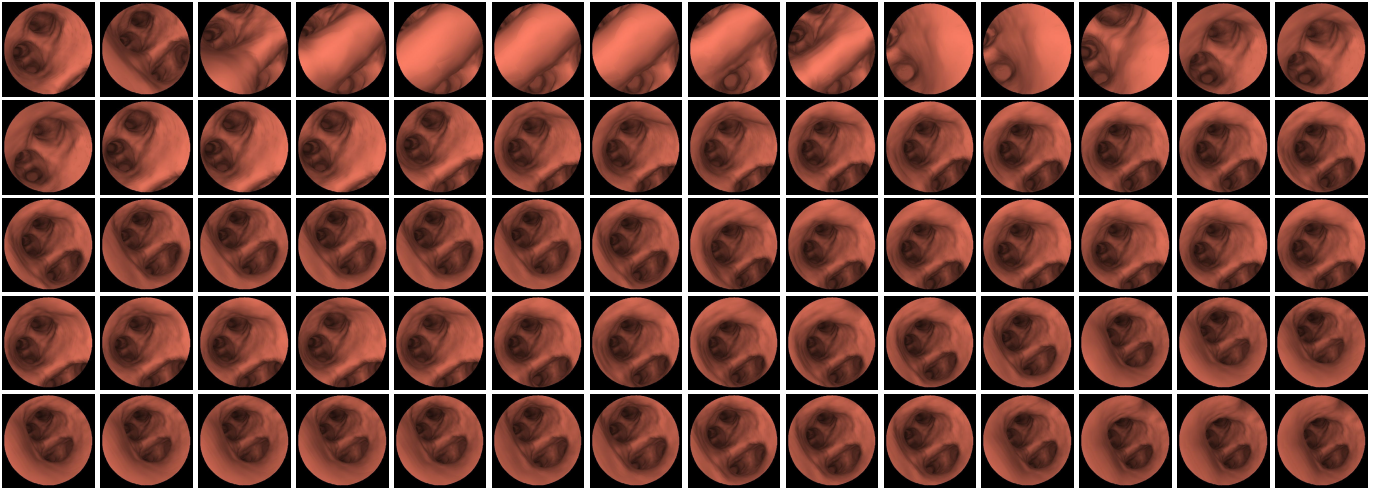




(a) Continuous bronchoscopic video frames 420~489 (*left*  $\rightarrow$  *right*, *top*  $\rightarrow$  *bottom*) of Experiment 7



(b) Corresponding virtual bronchoscopic images estimated from the method of Luo et al. (Luo et al., 2012b)



(c) Corresponding virtual bronchoscopic images estimated from our proposed OADE method

Figure 16: Example images from Experiment 7 illustrate that our proposed method can tackle respiratory motion and bronchoscopic image artifacts. Image artifacts, such as illumination changes (e.g., Frames 420, 421, and 422), collisions with bronchial walls (e.g., Frames 423~427), and jumps (e.g., Frames 428 and 49) occurred frequently. Our method still tracks bronchoscope motions successfully and greatly outperforms the method of Luo et al. (Luo et al., 2012b) since virtual images in (c) resemble video images much better than in (b).

rate adaptively, we proposed an improved DE method: OADE. With an application to bronchoscope motion estimation, our OADE algorithm can effectively fuse different sensory information of bronchoscopic video sequences, electromagnetic sensor measurements, and computed tomography images and provide more accurate and smoother bronchoscope tracking. The current tracking accuracy and smoothness were about 2.89 mm, 8.18° and 1.62 mm, 2.11°. The tracking performance was significantly improved, compared to state-of-the-art methods that at least show the tracking accuracy and smoothness of 3.96 mm, 10.47° and 4.08 mm, 2.65°. Future work includes the reduction of runtimes and clinical dataset validation.

## Acknowledgments

This work was partly supported by the Kakenhi “Modality-seamless navigation for endoscopic diagnosis and surgery assistance based on multi-modality image fusion” (25242047) funded by the Japan Society for the Promotion of Science (JSPS) and JSPS Bilateral International Collaboration Grant.

## References

- Brest, J., Greiner, S., Boskovic, B., Mernik, M., Zumer, V., 2006. Self-adapting control parameters in differential evolution: A comparative study on numerical benchmark problems. *IEEE Transactions on Evolutionary Computation* 10, 646–657.
- Chung, A.J., Deligianni, F., Shah, P., Wells, A., Yang, G.Z., 2006. Patient-specific bronchoscopy visualization through BRDF estimation and disocclusion correction. *IEEE Transactions on Medical Imaging* 25, 503–513.
- Dam, E.B., Koch, M., Lillholm, M., 1998. Quaternions, Interpolation and Animation. Technical Report DIKU-TR-98/5. University of Copenhagen, Denmark.
- Das, S., Suganthan, P.N., 2011. Differential evolution: A survey of the state-of-the-art. *IEEE Transactions on Evolutionary Computation* 15, 4–31.
- Deguchi, D., Mori, K., Feuerstein, M., Kitasaka, T., Maurer Jr., C.R., Suenaga, Y., Takabatake, H., Mori, M., Natori, H., 2009. Selective image similarity measure for bronchoscope tracking based on image registration. *Medical Image Analysis* 13, 621–633.
- Deligianni, F., Chung, A.J., Yang, G.Z., 2006. Nonrigid 2-D/3-D registration for patient specific bronchoscopy simulation with statistical shape modeling: Phantom validation. *IEEE Transactions on Medical Imaging* 25, 1462–1471.
- Doucet, A., Godsill, S., Andrieu, C., 2000. On sequential Monte Carlo sampling methods for Bayesian filtering. *Statistics and Computing* 10, 197–208.
- Feuerstein, M., Reichl, T., Vogel, J., Traub, J., Navab, N., 2009. Magneto-optical tracking of flexible laparoscopic ultrasound: Model-based online detection and correction of magnetic tracking errors. *IEEE Transactions on Medical Imaging* 28, 951–967.
- Gergel, I., dos Santos, T.R., Tetzlaff, R., Maier-Hein, L., Meinzer, H.P., Wegner, I., 2010. Particle filtering for respiratory motion compensation during navigated bronchoscopy, in: *Proc. SPIE MI 2010*, p. 76250W.
- Helferty, J.P., Sherbondy, A.J., Kiraly, A.P., Higgins, W.E., 2007. Computer-based system for the virtual-endoscopic guidance of bronchoscopy. *Computer Vision and Image Understanding* 108, 171–187.
- Isard, M., Blake, A., 1998. Condensation - conditional density propagation for visual tracking. *International Journal of Computer Vision* 29, 5–28.
- Levoy, M., 1990. A hybrid ray tracer for rendering polygon and volume data. *IEEE Computer Graphics and Applications* 10, 33–40.
- Liu, J., Lampinen, J., 2005. A fuzzy adaptive differential evolution algorithm. *Soft Computing - A Fusion of Foundations, Methodologies and Applications* 9, 448–462.
- Luo, X., Feuerstein, M., Deguchi, D., Kitasaka, T., Takabatake, H., Mori, K., 2012a. Development and comparison of new hybrid motion tracking for bronchoscopic navigation. *Medical Image Analysis* 16, 577–596.
- Luo, X., Feuerstein, M., Kitasaka, T., Imaizumi, K., Hasegawa, Y., Mori, K., 2010a. Towards hybrid bronchoscope tracking under respiratory motion: evaluation on a dynamic motion phantom, in: *Proc. SPIE MI 2010*, p. 76251B.
- Luo, X., Feuerstein, M., Kitasaka, T., Mori, K., 2012b. Robust bronchoscope motion tracking using sequential Monte Carlo methods in navigated bronchoscopy: dynamic phantom and patient validation. *International Journal for Computer Assisted Radiology and Surgery* 7, 371–387.
- Luo, X., Mori, K., 2012. Observation-driven adaptive differential evolution for robust bronchoscope 3-d motion tracking, in: *Proc. ACCV 2012*, pp. 259–271.
- Luo, X., Reichl, T., Feuerstein, M., Kitasaka, T., Mori, K., 2010b. Modified hybrid bronchoscope tracking based on sequential monte carlo sampler: Dynamic phantom validation, in: *Proc. ACCV 2010*, pp. 409–421.
- Mezura-Montes, E., Velazquez-Reyes, J., Coello Coello, C.A., 2006. Modified differential evolution for constrained optimization, in: *Proc. IEEE CEC 2006*, pp. 25–32.
- Mori, K., Deguchi, D., Akiyama, K., Kitasaka, T., Maurer Jr., C.R., Suenaga, Y., Takabatake, H., Mori, M., Natori, H., 2005. Hybrid bronchoscope tracking using a magnetic tracking sensor and image registration, in: *Proc. MIC-CAI 2005*, pp. 543–550.
- Price, K.V., Storn, R.M., Lampinen, J.A., 2005. Differential evolution: A practical approach to global optimization. Springer-Verlag, New York.
- Qin, A.K., Huang, V.L., Suganthan, P.N., 2009. Differential evolution algorithm with strategy adaptation for global numerical optimization. *IEEE Transactions on Evolutionary Computation* 13, 398–417.
- Schwarz, Y., Greif, J., Becker, H.D., Ernst, A., Mehta, A., 2006. Real-time electromagnetic navigation bronchoscopy to peripheral lung lesions using overlaid CT images: the first human study. *Chest* 129, 988–994.
- Soper, T.D., Haynor, D.R., Glenny, R.W., Seibel, E.J., 2010. In vivo validation of a hybrid tracking system for navigation of an ultrathin bronchoscope within peripheral airways. *IEEE Transactions on Biomedical Engineering* 57, 736–745.
- Storn, R., Price, K.V., 1997. Differential evolution - A simple and efficient heuristic for global optimization over continuous spaces. *Journal of Global Optimization* 11, 341–359.
- Teo, J., 2006. Exploring dynamic self-adaptive populations in differential evolution. *Soft Computing - A Fusion of Foundations, Methodologies and Applications* 10, 673–686.
- Tiede, U., Hoehne, K.H., Bomans, M., Pommert, A., Riemer, M., Wiebecke, G., 1990. Investigation of medical 3D-rendering algorithms. *IEEE Computer Graphics and Applications* 10, 41–53.
- Wang, Z., Bovik, A.C., 2002. A universal image quality index. *IEEE Signal Processing Letters* 9, 81–84.
- Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P., 2004. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing* 13, 600–612.
- Zhang, J., Sanderson, A.C., 2009. JADE: Adaptive differential evolution with optional external archive. *IEEE Transactions on Evolutionary Computation* 13, 945–958.