The Concept of Evolutionary Computing for Robust Surgical Endoscope Tracking and Navigation

Ying Wan

School of Computing and Communications
University of Technology, Sydney

This dissertation is submitted for the degree of
Doctor of Philosophy in Computer Science

Faculty of Engineering and
Information Technology

April 2017

I would like to dedicate this thesis to my remarkable husband, my adorable son, my loving parents, and the memory of my grandmother who just passed away December, 2015.

Declaration

I hereby declare that except specific references, the content of this dissertation is original and has not been submitted in whole or in part for consideration for any other degree or qualification from any other universities. This dissertation contains nothing which is the outcome of work that has been done in collaboration with others, except as specified in the text and acknowledgments. In general, this dissertation totally consists of about 65,000 words including appendices, bibliography, footnotes, tables, equations, and 150 figures.

Ying Wan
April 2017

Acknowledgements

I am using this opportunity to express my gratitude to thank Professors Xiangjian He and Xiongbiao Luo who are my principle and co-supervisors, respectively, for their many suggestions, constant encouragement, and support during my PhD candidature. Moreover, I am sincerely grateful to them for sharing their truthful and illuminating vision on my work and have the honor of studying and working with them in the past three years and six months.

I certainly appreciate the support of the International Research Scholarship (IRS) provided by the Faculty of Engineering and Information Technology (FEIT), University of Technology, Sydney (UTS). I absolutely appreciate the financial support for my living in the PhD study.

I wish to thank my fellow colleagues and the staffs of the faculty for providing various assistance for the completion of this research work. In particular, Qiang Wu, Wenjing Jia, Min Xu, Sheng Wang, Minqi Li, Tao Zeng, and David Du, for their invaluable supports.

Last but not least, I would like to thank my husband and my parents for their patient, understanding, and wide support. This PhD thesis could not have been available without their encouragements and financial assistance during the past three years and six months.

Abstract

Navigated endoscopy is generally agreed to be the next generation of interventional or surgical endoscopy. It usually combines pre- and intra-operative imaging information to guide physicians during endoscopic procedures. However, endoscope three-dimensional motion tracking that spatially and temporally synchronizes various sensory information still remains challenging for developing different endoscopic navigation systems. To navigate or track the surgical endoscope, three modalities of sensory information are utilized in endoscopic procedures: (1) preoperative images, i.e., three-dimensional CT images, (2) two-dimensional video sequences from the endoscopic camera, and (3) location measurements, attaching an electromagnetic sensor at the endoscope distal tip for measuring the temporal endoscope movement. In this respect, endoscope tracking and navigation aims to fuse these various modalities information to accurately and robustly locate or fly through the endoscope at any interest of regions. Unfortunately, fusing the multimodal information is still an open issue due to the information incompleteness, e.g., image artifacts, tissue deformation, and sensor output inaccuracy in computer assisted endoscopic interventions.

This thesis work focuses on fusing the multimodal information for accurate and robust endoscope tracking and navigation. A novel framework of multimodal information fusion is proposed to use evolutionary computing for endoscopic navigation systems. Several main contributions of this dissertation are clarified as follows. First, the concept of evolutionary computing was initially introduced to assist minimally invasive endoscopic surgery. Next, this work modified two evolutionary algorithms of particle swarm optimizer and differential evolution and proposed an enhanced particle swarm optimizer (EPSO) and observation-driven adaptive differential evolution (OADE). EPSO can adaptively update evolutionary parameters in accordance with spatial constraints and the current observation. OADE performs 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. Additionally, the improved evolutionary computing algorithms are applicable to computer vision tasks, e.g., object tracking, motion estimation, and stochastic optimization.

The experimental results demonstrate that the proposed evolutionarily computed endoscopic tracking and navigation approaches in this dissertation provide a more accurate and robust endoscopic guidance framework than state-of-the-art methods. Based static phantom data validation, the average guidance accuracy of the EPSO framework was about 3.0 mm, its average position smoothness was 1.0 mm, and its average visual quality was improved to 0.29. By evaluating on a dynamic phantom, the OADE 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.

In conclusion, the concept of evolutionary computation is a promising strategy to improve endoscopic tracking and navigation for minimally invasive surgery. The validation demonstrated its effectiveness to improve the guidance accuracy, visual quality, and tracking smoothness during endoscopic surgery. Future work includes surgical data validation, real-time processing, and translation to clinical applications.

Keywords

Minimally invasive surgery — Bronchoscope — Endoscopy — Endoscopic navigation — Endoscopic video processing — Computer-assisted intervention — 2D/3D registration — Pre- and intra-operative imaging — Electromagnetic tracking — Evolutionary computing — Particle swarm optimizer — Differential evolution

Table of contents

List of figures

List of tables

Chapter 1

Minimally invasive endoscopic surgery

Lung cancer is the leading cause of cancer death worldwide with an estimated 1.8 million new cases and 1.6 million deaths in 2012 [10]. Early detection and treatment are significant to reduce the rate of incidence and death. Physicians usually implement a two-stage procedure to detect lung cancer: (1) preoperative imaging examination – a noninvasive specialized three-dimensional (3D) scan of a patient chest, e.g., using computed tomography (CT) and magnetic resonance imaging (MRI) scanners to initially inspect pulmonary diseases or abnormalities, and (2) intraoperative (or interventional) bronchoscopy – a minimally invasive medical procedure that uses a bronchoscope to insert the body through natural orifices (e.g., mouth and nose) and directly visualizes inside the airway tree and performs interventional pulmonary procedures, e.g., transbronchial needle aspiration (TBNA) and transbronchial lung biopsy (TBLB), to sample cancerous or suspicious tissue.

Interventional bronchoscopy plays an important role in lung cancer diagnosis, staging, and treatment. In such an endoscopic procedure, the flexible bronchoscope that is usually integrated with a video camera at its distal tip provides two-dimensional (2D) video images that show the intrinsic structure of the bronchial tree. Figure 1 shows examples of one CT slice image and one bronchoscopic video image. Based on 3D preoperative images (i.e., CT slices) and temporal 2D video sequences, the physician mentally registers these 2D images to the 3D CT slice space for determining where the bronchoscope is flying through the anatomical structure of the bronchial tree. After the bronchoscope reaches the target regions or tumors, the physician performs biopsy or treatment. During endoscopic interventions, the key question lies in accurately and temporally determining the endoscope pose information with six degrees of freedom (6DoF) position and orientation parameters in a reference coordinate system, i.e., without accurate localization of the bronchoscope, it is difficult to guarantee direct position of target regions (e.g., bronchial lesions, pulmonary nodules, or lymph nodes). Unfortunately, the endoscope itself only provides 2D images without any depth

information during conventional endoscopy. Furthermore, the bronchial tree is generally a complex anatomical structure with thousands of branches. Such complication leads to locate the endoscope at target regions with much difficulty. In addition, the physician's mental fusion of 3D CT and video images in the operating room possibly results in excessive cognitive loading and surgical error or risk.

Fortunately, with recent technological advances in augmented reality (AR), real-time endoscopic imaging, and miniaturized transducer or sensor techniques, endoscopic navigation systems with endoscope 3D motion tracking are increasingly developed to solve clinical issues during endoscopic interventions. By combining conventional endoscopy (without any navigation information) with different current AR and tracking techniques (e.g.,volume rendering and electromagnetic trackers), the development of navigated endoscopy aims to realize two fundamental functions (1) localization and (2) navigation, which correspond to two basic questions from physicians in operating rooms: (1) where to go and (2) how to go. Furthermore, real-time visualization for various modalities, e.g., pre-operative CT slices, ultrasound images, VB renderings, and endoscopic video images, is another main benefit of navigated endoscopy; this yield enables physicians to visually and interactively manipulate or control the endoscope and its accessories during endoscopic interventions. Additionally, it can be used for 3D path planning to generate the optimal route for passage of endoscopic instruments and can further manage or backup the motion trajectories of endoscopes or medical tools. In general, endoscope tracking and navigation (ENT) systems enable physicians to accurately localize target regions, correctly sample suspicious tissues, and avoid unnecessary surgical interventions in diagnosis, staging and treatment of pulmonary diseases at the early stage.

1.1 Minimally invasive surgery

Minimally-invasive surgeries have been enabled by the advance of various medical technologies. Surgery by definition is invasive and many operations requiring incisions of some size, are referred to as open surgery. Incisions made can sometimes leave large wounds that are painful and take a long time to heal. Minimally-invasive surgery refers to surgical techniques that limit the size of incisions needed and so lessens wound healing time, associated pain and risk of infection. This type of surgery is usually performed using thin-needles and an endoscope to visually guide the surgery. The goal of minimally invasive surgery is to reduce postoperative pain and blood loss, speed recovery, and lessen scarring.

1.2 Surgical endoscopy

Surgical endoscopy is a group of minimally invasive surgical techniques, My study specifically focus on surgical bronchoscopy. Bronchoscopy is an endoscopic technique of visualizing the inside of the airways for diagnostic and therapeutic purposes. An instrument (bronchoscope) is inserted into the airways, usually through the nose or mouth, or occasionally through a tracheostomy. This allows the practitioner to examine the patient's airways for abnormalities such as foreign bodies, bleeding, tumors, or inflammation and take specimens from inside the lungs. The construction of bronchoscopes ranges from rigid metal tubes with attached lighting devices to flexible optical fiber instruments with realtime video equipment.

Currently, the purpose of interventional bronchoscopy basically consist of two parts: (1) diagnosis and staging, and (2) therapy and treatment. Different goals correspond to different bronchoscopic procedures where different diagnostic and therapeutic techniques or ancillary tools are utilized during intra-bronchoscopy. To realize these different roles, physicians usually perform various diagnostic and therapeutic ancillary tools or manipulations during bronchoscopic interventions in operation rooms. However, it is the bronchoscope that is used in applications of ancillary tools such as transbronchial lung biopsy and balloon dilatation. Therefore, the success of these roles depends heavily on the position and orientation tracking of the bronchoscope. However, it is necessary to first introduce the different bronchoscopic procedures with different diagnostic and therapeutic tools

1.2.1 Diagnosis and Staging

The basic diagnostic role of interventional bronchoscopy is bronchoscopic visualization to examine endobronchial abnormalities such as possible tumors, obstructions, secretions or foreign bodies [11].

Endobronchial ultrasonography (EBUS) is a diagnostic modality in which a miniature ultrasonic probe is introduced into the tracheobronchial lumen to provide the sonographic images of the peribronchial tissue. Radial probe EBUS is most commonly used for localization of peripheral lung nodule before obtaining tissue sample during bronchoscopy. These multiple diagnostic techniques mainly consist of the following manipulations or ancillary tools during bronchoscopic interventions [12]: (1) bronchial brushing, (2) bronchial washing, (3) bronchoalveolar lavage (BAL), (4) endobronchial biopsy, (5) transbronchial lung biopsy (TBLB), and (6) transbronchial needle aspiration (TBNA). These tools and their combinations play different roles during the assessment of pulmonary diseases.

Staging is the process of determining how much cancer there is in the body and where it is located. The treatment options largely depend on the stage of the disease, so Staging of lung

(a) Cytology brushes (b) Forceps biopsy (c) Aspiration biopsy needles

Figure 1.1: Examples of diagnostic and therapeutic ancillary tools usually used during interventional bronchoscopy (Images courtesy of Olympus, Europe).

cancer is of paramount importance. Bronchoscopy is utilized in sampling of the respiration tract secretions and cells, and biopsy of the airway, lung, and mediastinal structures. Since the introduction of transbronchial needle aspiration (TBNA) in flexible bronchoscopy in 1983, conventional TBNA (has been technically well-established and expanded its role in diagnosis and staging of lung cancer. [13]

1.2.2 Therapy and Treatment

Interventional bronchoscopy is an evolving field that utilizes minimally invasive modalities for therapy and treatment of suspected lung cancers. Various bronchoscopic therapeutic procedures performed in collaboration with surgery, external beam radiation, or chemotherapy mainly consist of: (1) laser therapy, (2) electrocautery and cryotherapy, and (3) stent placement. These techniques remove or relieve central airway obstructions or stenoses resulting from various malignant and benign airway lesions.

Laser therapy provides a common routine therapeutic tool to treat lung tumors in a relatively safe way, as other tools (e.g., forceps biopsy) have dangerous risks (e.g., fatal bleeding) during interventions. It is mainly used to alleviate the symptoms of airway obstruction, particularly in patients whose large airways are blocked by primary thoracic or metastatic malignant tumors. Various beneficial effects of laser therapy were reported in many clinical applications. Treatment success rates in selected patients is based on the experience and training of bronchoscopists, and the therapy has also been demonstrated to improve quality of life and functional status, and, in some cases, to extend survival [14].

Lasers can be used in 2 ways to treat cancer: (1)To shrink or destroy a tumor with heat (2)To activate a chemical –known as a photosensitizing agent – that kills only the cancer cells.Laser treatment normally used with other cancer treatments, such as chemotherapy or radiation.

Electrocautery was developed to address the limitations of laser therapy such as its high cost, relative shortage of accessibility, and its constraints on rigid bronchoscope. It is usually applied to deal with benign and malignant lesions by using a probe to puncture tissues (unlike laser therapy works in a non-contact means). It has proved to be as effective as laser therapy. However, the penetration depth and resulting injury from electrocautery are much more difficult to control than with laser therapy. Cryotherapy is designed to destroy pathologic tissue that is located at poorly cellular regions or those with limited vascularization, such as benign fibrotic structures and lipomas. Compared to laser therapy or electrocautery, its main advantages are safety, user-friendliness, and cost-effectiveness, particularly as it has no any risk of endobronchial fire; but is more difficult to control the penetration depth.

Stent placement is another different approach to alleviate the airway obstruction symptoms caused by either a malignant or benign obstruction that has failed other medical, surgical, or bronchoscopic therapies. It enables sustainable palliation for progressive symptoms of airway obstruction in patients. However, stents can only be considered as mechanical palliation but not as a cure for potential bronchial diseases.[14]

1.3 Navigated Bronchoscopy Overview

Minimally invasive bronchoscopic procedures or interventions are often highly interactive; various operations are performed immediately by physicians in surgical rooms. While executing these operations or decisions, physicians expect continuously visual feedback , to assist them to control medical tool movements. Such visual feedback requires the estimation and tracking of position and orientation of endoscopes or other medical tools in a 3D space. In navigated bronchoscopy, motion tracking usually denotes the spatial synchronization of pre-operative and intra-operative imaging data to identify and understand anatomical structures. We develop navigated bronchoscopy to enhance the capability of medical doctors to diagnose and treat patients, for example, by enabling them to rapidly and accurately localize the bronchoscope tip to tumor regions for a biopsy.

Most navigated bronchoscopy systems rely mainly on various imaging modalities, computer vision and graphics techniques, and medical image analysis. These techniques make up the foundation of the field of navigated bronchoscopy. In general, a bronchoscopic navigation system can simultaneously incorporate anatomic imaging in different modalities, endoscope motion tracking, as well as real-time visualization techniques.[14]

1.4 Related research field

Previous sections stated the benefit of surgical endoscope or relation of this thesis. However, this thesis work is strongly related to other technical fields, especially augmented reality (AR), real-time endoscopic imaging, and miniaturized transducer or sensor techniques. Endoscope with digital cameras provides an opportunity to apply computer vision techniques to offer visualization augmentations, for example, structure reconstruction, image registration, and motion recovery or tracking. As camera motion tracking or pose (position and orientation) estimation is a very interesting and challenging topic in the aspect of computer vision, this work is to combine conventional endoscopy (without any navigation information) with different current AR and tracking techniques. Addressing the problem of endoscopic camera motion tracking will contribute to the field of computer vision and also extends to different computer vision technique applications in medical community.

1.5 Dissertation organization

This dissertation is motivated by how to accurately and robustly localize the bronchoscope or medical tools and track their motion, and to construct bronchoscopy navigation systems to assist different bronchoscopic techniques or procedures to be performed during bronchoscopic interventions. In this dissertation, the proposed approaches to estimate the bronchoscope movement to determine its position and orientation inside bronchi may lead to improved diagnostic yield from bronchoscopic procedures. This dissertation is comprised of six chapters.

This chapter discusses a brief history and background, basic roles of bronchoscopy, current status of minimally invasive diagnostic and therapeutic bronchoscopy and the advantages and related research field.

Chapter 2 firstly reviews state-of-the-art developments in endoscope tracking and navigation for constructing the next generation of endoscopy systems, followed by challenges that remain in current available endoscopic navigation techniques.

Continuously, the following three chapters are the main body of this thesis. They present three proposed methods of medical instruments motion tracking for bronchoscopic navigation. Chapter 3 introduces a registration method betweenthe electromagnetic tracker (or sensor) and pre-clinical images. Multiple points from sensor measurements and endoscope tip center is generated by a adaptive strategy which is introduced in this article. From these generated points, we adaptively choose the optimal point, which is the closest to its assigned the centerline of the hollow organ, to perform registration. Chapter 4 proposes

a new framework on the basis of an enhanced particle swarm optimization method based on particle swarm optimization method to effectively fuse these information for accurate and continuous endoscope localization that can tackle the problem that evolutionary computation method usually limits its possible premature convergence and evolutionary factors. Furthermore, in Chapter 5 proposes an observation-driven adaptive differential evolution algorithm based on differential evolution algorithm that fuses bronchoscopic video sequences, electromagnetic sensor measurements, and computed tomography images for accurate and smooth bronchoscope three-dimensional motion tracking

Finally, Chapter 6 concludes this thesis work on medical tool localization and tracking for navigation and provides a summary of the accomplishments discussed in the dissertation. several potential limitation and promising directions are discussed.

Chapter 2

Endoscope tracking and navigation

2.1 State of the art

Endoscopic interventions are widely performed for diagnosis and treatment of different cancers, e.g., bronchus and lung cancers via bronchoscopic surgery [15–18], bowel cancer by colonoscopic surgery [19], brain tumors using neurosurgery [20, 21]. Guided endoscopic procedure is generally recognized as the next generation of interventional endoscopy. Such a guidance usually involves different modalities of sensory information: (1) pre-operative information, e.g., 3D CT images, (2) 2D endoscopic video sequences, and (3) positional sensor measurements, e.g., using an electromagnetic (EM) tracking system with a 6DoF sensor attached at the endoscope tip (*Northern Digital Inc.*, Waterloo, Canada)[1]. However, only using one of three kinds of information is very difficult to accurately navigate or localize the endoscope to cancerous regions where interventions (e.g., biopsy and resection) must be performed. At this point, the endoscopic guidance is considered as a procedure of multimodal information fusion.

Endoscope tracking and navigation (ENT) is developed to solve the problem of the endoscope tip location during endoscopic intervention on the basis of the multimodal information above. Various methods to navigate or track the endoscope have been published in the literature. They are divided into two major categories (or a combination of both): (1) CT-video registration (CVR) and (2) electromagnetic tracking (EMT).

The CVR methods usually define an optimization function to minimize the pixel differences between endoscopic video images and 2D virtual images generated from pre-operative 3D CT images by volume or surface rendering techniques [22, 23]. Helferty et al. reported an image registration method on the basis of normalized mutual information (NMI) for

[1] http://www.ndigital.com/

endoscope tracking [17]. Chung et al. registered video and virtual images to estimate the endoscope motion by recovering a bidirectional reflectance distribution function (BRDF) [16]. 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 endoscope tip [24], which method was further improved by combining the scale invariant feature transform method [18]. Figure 2.1 shows an example of CVR-based endoscopic navigation system interface. The EMT approaches use an electromagnetic tracker (EMT) with a position sensor fixed at the bronchoscope tip to directly measure endoscope motion [4]. Mori et al. proposed a hybrid method that combines CT-video registration and EMT to track the endoscope [5], which method was further improved by using Kalman filtering and the condensation algorithm [25, 8, 3]. Gergel et al. used particle filtering to compensate for tissue deformation due to breathing motion [26, 27]. 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 [28]. Magnetic field distortion is very difficult to correct without optical tracking [29]. Although both currently available CVR and EMT navigation methods work well, there are still several challenges for endoscope tracking and navigation during image-guided endoscopic interventions.

2.2 Current problems and challenge

ENT is generally a procedure of multimodal information fusion. Unfortunately, it still remains challenging to effectively and efficiently integrate three modalities of sensory information for endoscopic guidance since the sensory information is incomplete or uncertainty that means their limitations:

- **Tissue deformation**. Preoperative image information (e.g., CT slices) before endoscopic surgery cannot record any information about patient movements (e.g., respiration) which are exactly appeared in the operating room.

- **Endoscopic image artifacts**. Inter-reflection, motion blurring and bubbles unavoidably occur in endoscopic videos, resulting in a large number of low-quality images. In particular, CVR methods rely heavily on the video image quality that might be degenerated by image artifacts and also easily gets trapped in the local minima in optimization without a proper initial estimate.

- **Inaccurate sensor measurement**. The positional EM sensor's accuracy can be heavily deteriorated by patient movements, e.g., breathing motion and cough.

- **Jitter sensor outputs**. Jitter error happens frequently due to the magnetic field distortion that is caused by ferrous metals or conductive material within or close to the working volume of EM tracking systems. Due to the inaccurate and jitter sensor measurements, current commercially available endoscope tracking and navigation systems (e.g., superDimension), which are increasingly used in clinical applications [4, 30, 31], were reported their diagnostic accuracy between 59% and 74%, without sensitivity to tumor size [32, 33].

The goal of this work is to address these above limitations by evolutionarily computing the multimodal information to construct accurate and robust endoscope tracking and navigation systems for endoscopic interventions in operating rooms.

2.3 New concept: Evolutionary computing

Evolutionary computing is an active research field of artificial intelligence in computer science. It simulates biological evolution behaviors (e.g., reproduction, mutation, recombination or crossover, parent and survivor selections) to solve computational problems. Evolutionary algorithms are generally defined as a population-based stochastic optimization approach. Based on these biological behaviors, evolutionary procedures are activated by two sources or forces [34]:

- **Propagation**: Using various operators (e.g., mutation and crossover) as the force to diversify the population
- **Selection**: Employing the quality metric or function as the force to determine the survivor in the population

The population propagation and selection is integrated to optimize the evolution processing and find the optimal solution.

An evolutionary algorithm basically contains several important elements (Figure 2.4), as discussed as follows [34, 35]:

- **Population**. The population is defined as a number of individuals or particles that represent potential solutions for computational problem. Before the first iteration during optimization, each individual or particle in the population should be initialized with predefined values.
- **Operations**. The prorogation step with different operations on all the individuals aims to update the population to new states and approximate the optimal solution in a

searching space. On the other hand, the prorogation also increases the diversity of the population. The more diverse of the population, the more potential solutions can be obtained, and stronger exploration capacity, which indicates sufficiently covering the solution space to find the global optimum, can be achieved.

- **Evaluation**. Each propagated individual or particle is evaluated by an assessment function that is typically defined a quality measure to compute the individual's fitness value. Such a value denotes the individual's exploitation capacity that indicates sufficient fitness pressure to obtain even better solutions from good ones in the searching space. Basically, the higher the individual fitness, the more powerful individual exploitation.

- **Selection**. Based on all the fitness values of the evaluated population, the selection stage is to compare the propagated individual in the current generation to the old individual in the previous generation and survive the particles with the better quality or fitness value.

- **Termination**. The stopping condition is usually predefined for every evolutionary algorithm. Otherwise, evolutionary computing gets trapped in an endless loop and never stopping. If the computational problem has a known optimal fitness value, approximating this fitness value is naturally considered as the termination condition. Unfortunately, in most cases, it is difficult to know such an optimal fitness of the computational problem. Hence, fixing the number of generations in evolutionary computing is an easy way to stop the algorithm.

In general, evolutionary algorithms are widely applied to computational and optimization problems in various research fields, e.g., segmentation, visual tracking, video surveillance, and manufacturing process planning [34, 36]. In this thesis work, an accurate and robust multimodal data fusion on the basis of evolutionary algorithms with enhanced particle swarm optimization (EPSO) and observation-driven differential evolution (ODDE) is proposed for endoscopic navigation.

Before using evolutionary computing for endoscopic navigation, multimodal data involved in ENT are briefly reviewed.

2.4 Definition and formulation

This section defines the problem of endoscope tracking and navigation and parametrizes the endoscope 3D movement.

2.4.1 Endoscopic camera geometry

The camera geometry establishes the connection between the image plane and the world space of objects [37]. It basically involves with three coordinate systems of image, camera, and the world (Figure 2.2). The spatial relationships among the image, the camera, and the world can be represented by two projection matrices, i.e., camera intrinsic matrix $\mathbf{K}$ (camera→image) and extrinsic matrix $\mathbf{T}$ (world→camera).

Let $\mathbf{q}$, $\mathbf{Q}_C$, and $\mathbf{Q}_W$ be the points at the image, the camera, and the world spaces, respectively, then coming to the following equation:

$$\mathbf{q} = \mathbf{K}\underbrace{\mathbf{T}\mathbf{Q}_W}_{\mathbf{Q}_C} = \mathbf{K}\underbrace{\begin{pmatrix} {}^{C}\mathbf{R}_W & {}^{C}\mathbf{t}_W \\ 0 & 1 \end{pmatrix}\mathbf{Q}_W}_{\mathbf{Q}_C}, \tag{2.1}$$

where translation ${}^{C}\mathbf{t}_W = ({}^{C}t_W^x, {}^{C}t_W^y, {}^{C}t_W^z)^T$ (the camera center position in the world) and rotation matrix ${}^{C}\mathbf{R}_W$ is represented

$${}^{C}\mathbf{R}_W = \begin{pmatrix} C_3C_2 & C_3S_2S_1 - S_3C_1 & C_3S_2C_1 + S_3S_1 \\ S_3C_2 & S_3S_2S_1 + C_3C_1 & S_3S_2C_1 - C_3S_1 \\ -S_2 & C_2S_1 & C_2C_1 \end{pmatrix}, \tag{2.2}$$

where $S_1 = \sin\alpha$, $S_2 = \sin\beta$, $S_3 = \sin\gamma$, $C_1 = \cos\alpha$, $C_2 = \cos\beta$, $C_3 = \cos\gamma$, and α, β, and γ denote three Euler angles.

2.4.2 Parameterization

Navigated endoscopy involves multimodal information corresponding to various coordinate systems where the endoscope or surgical tool is placed. It includes generally three coordinate systems: (1) world/patient, (2) endoscopic camera, and (3) CT data (Figure 2.3). pre-operative imaging data. Endoscope tracking and navigation aims to synchronize these coordinate systems, i.e., addressing the subject of how these coordinate systems are associated with each other from different transformation matrices ${}^{CT}\mathbf{M}_C$, ${}^{CT}\mathbf{M}_W$, and ${}^{W}\mathbf{M}_C$, and eventually determining the endoscopic camera's 6DoF position and orientation in the 3D CT image coordinate system:

$${}^{CT}\mathbf{M}_C = {}^{CT}\mathbf{M}_W {}^{W}\mathbf{M}_C, \tag{2.3}$$

If the world coordinate system is defined as the EM tracking coordinate system including the positional EM sensor coordinate system S (Figure 2.6), Equation 2.3 becomes:

$$^{CT}\mathbf{M}_C = ^{CT}\mathbf{M}_{EM}\,^{EM}\mathbf{M}_S\,^{S}\mathbf{M}_C, \tag{2.4}$$

where $^{S}\mathbf{M}_C$ is computed by hand-eye calibration methods, $^{CT}\mathbf{M}_{EM}$ is determined by mark-free or marker-based registration approaches, and the sensor measurement $^{EM}\mathbf{M}_S$ with position and orientation of the EM tracking system.

In general, endoscope tracking and navigation seeks to estimate transformation $^{CT}\mathbf{M}_C$ with position and rotation from the endoscopic camera to the 3D CT image coordinate system.

2.5 Multimodal information fusion

Multimodal information in ENT was illustrated in Figure 2.5.

- **CT slices**. Before performing endoscopic examinations and interventions, physicians usually use 3-D imaging devices such as CT scanners to acquire pre-operative images.The CT images are employed to generate 2D virtual images by volume rendering techniques [38].

- **Endoscopic videos**. The endoscopic camera is directly inserted to observe organ interior structures. Endoscopic images display sufficient organ interior surface information. Due to 2D video images without depth information to target regions, integrating CT with 2D endoscopic images can guide endoscopic procedures.

- **Sensor measurements**. Ultra-miniature sensors, i.e, EM sensors. which are externally attached to the endoscope's distal tip surface or its working channel, are increasingly used to measure the endoscopic camera motion. However, the sensor measurements might be inaccurate due to tissue deformation, e.g., respiration.

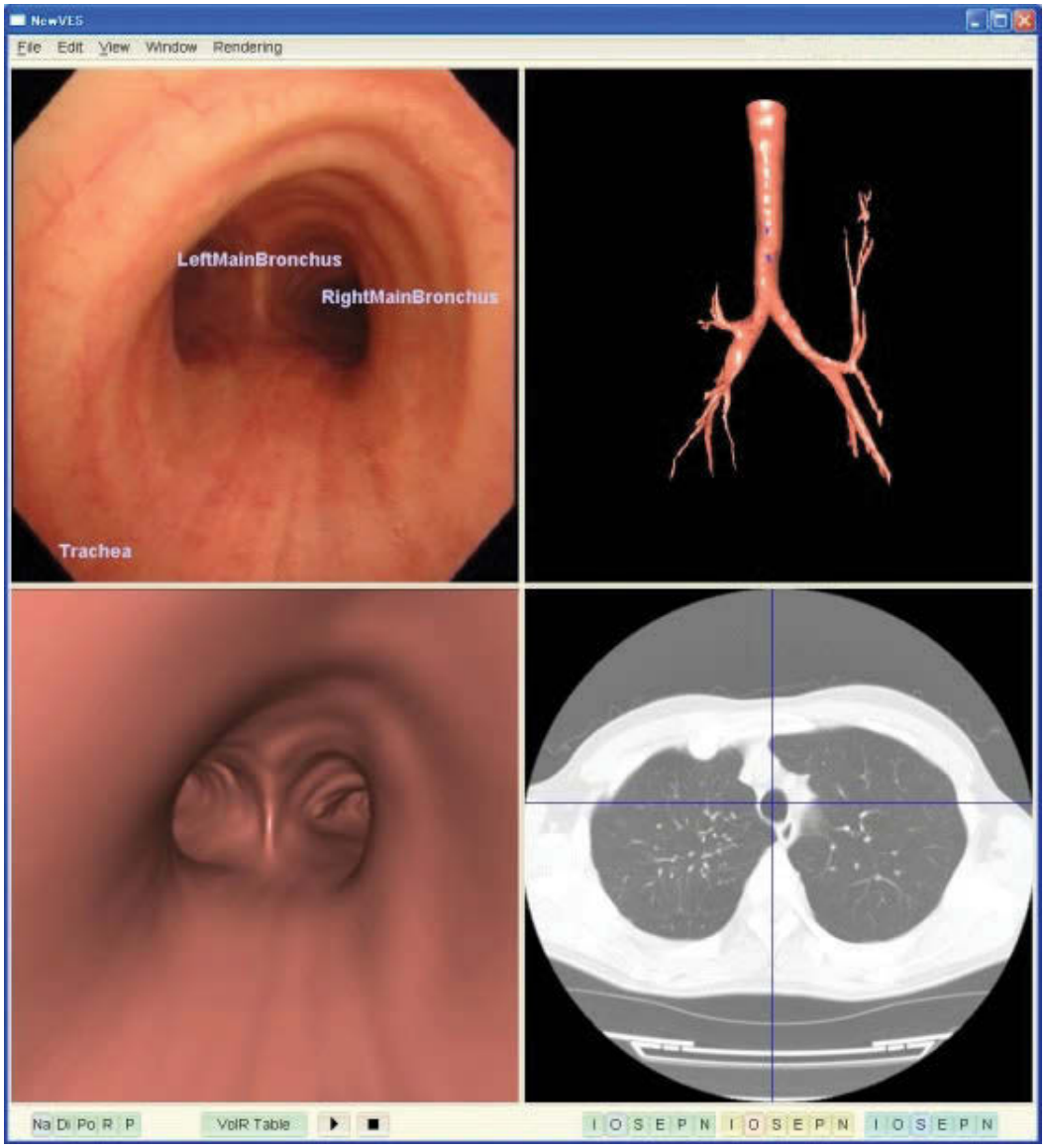


Figure 2.1: Composite view of an endoscope tracking and navigation system that contains four windows providing with useful information: Endoscopic video sequences (upper left), 2D virtual rendering images (lower left) generated from previously acquired CT data and corresponding to video images, a bird's view (upper right) displaying the endoscope trajectory, and a slice view (lower right) showing an axial slice, in which a cross cursor marks the current endoscope location. All windows were synchronized to provide critical visual and structural information during endoscopic intervention.

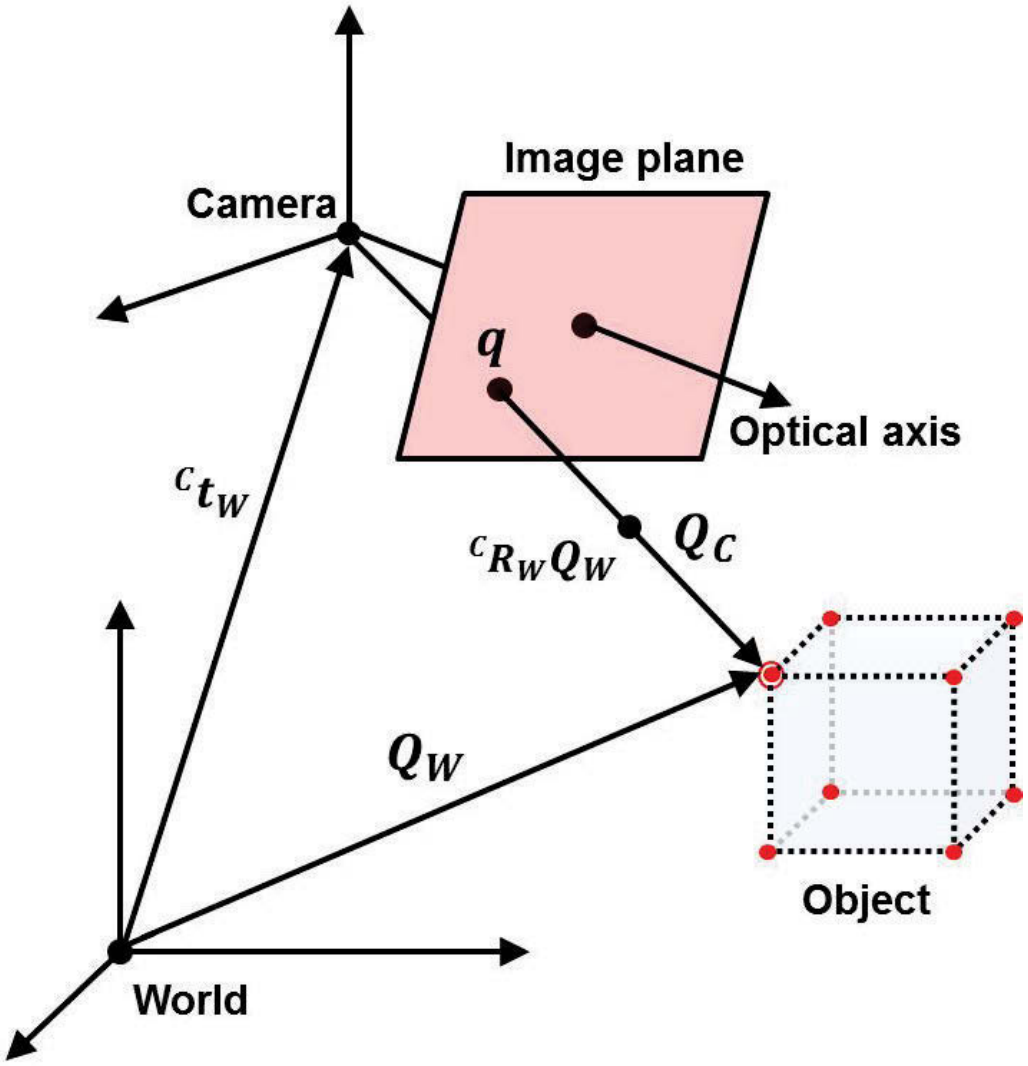


Figure 2.2: Camera geometry with three coordinate systems

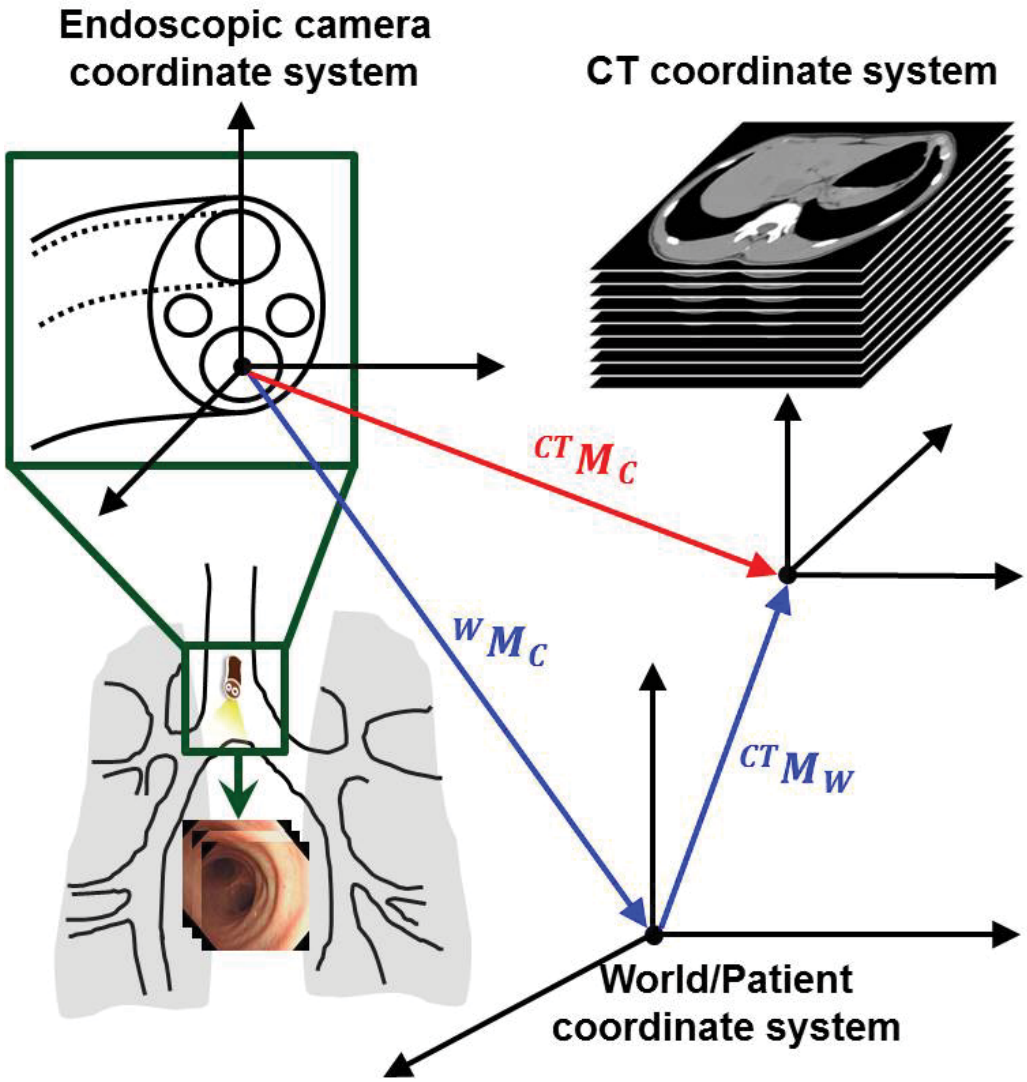


Figure 2.3: Various spatial transformations among different coordinate systems involved in endoscopic navigation systems

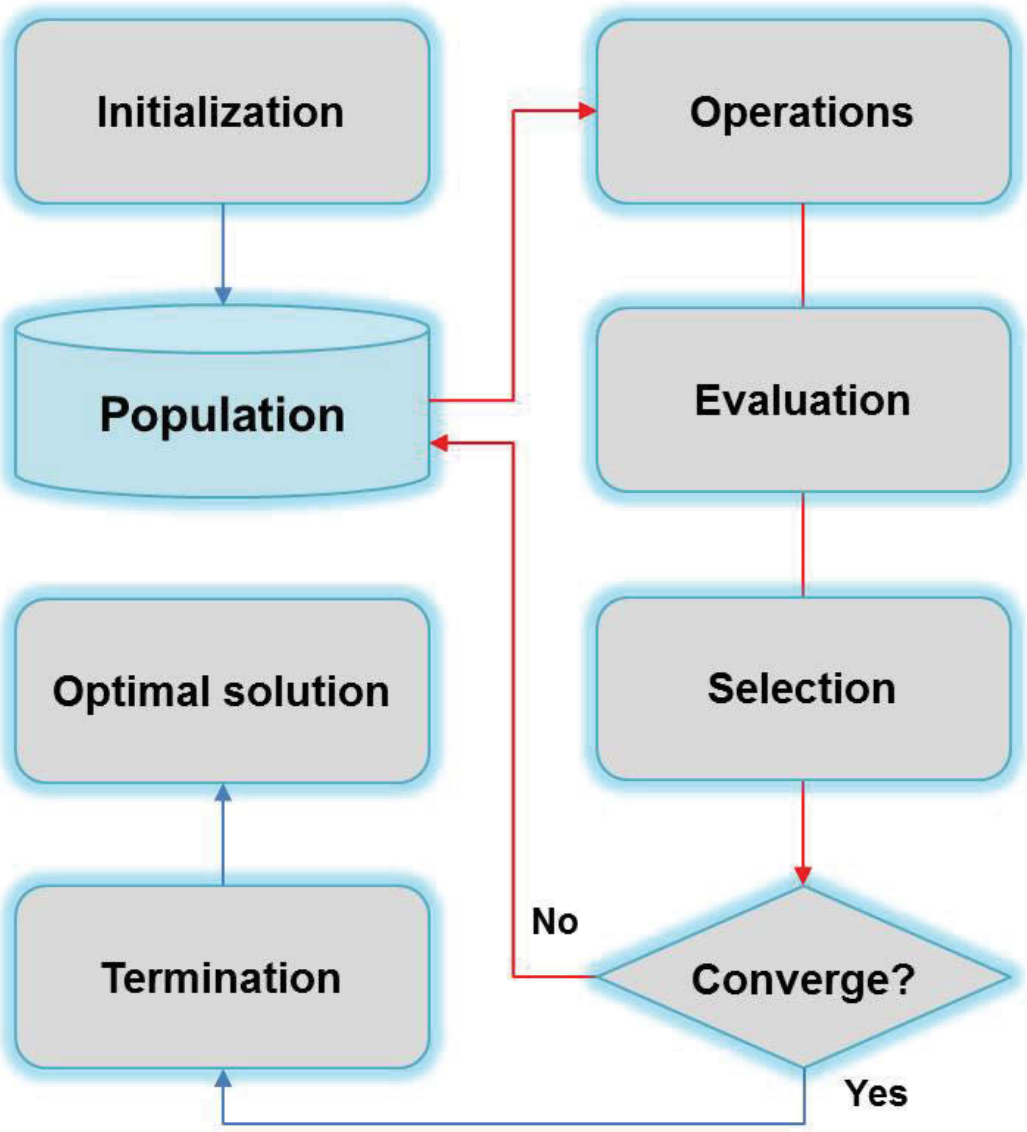


Figure 2.4: General flowchart of evolutionary computing

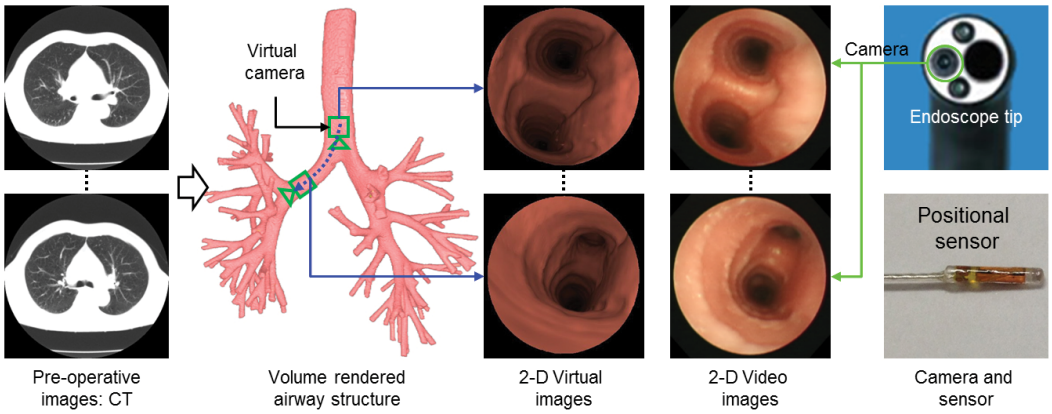


Figure 2.5: Multimodal information including the CT images used to generate 2D virtual rendering images, endoscopic video images, and positional sensor measurements involved in an electromagnetically guided endoscopic procedure

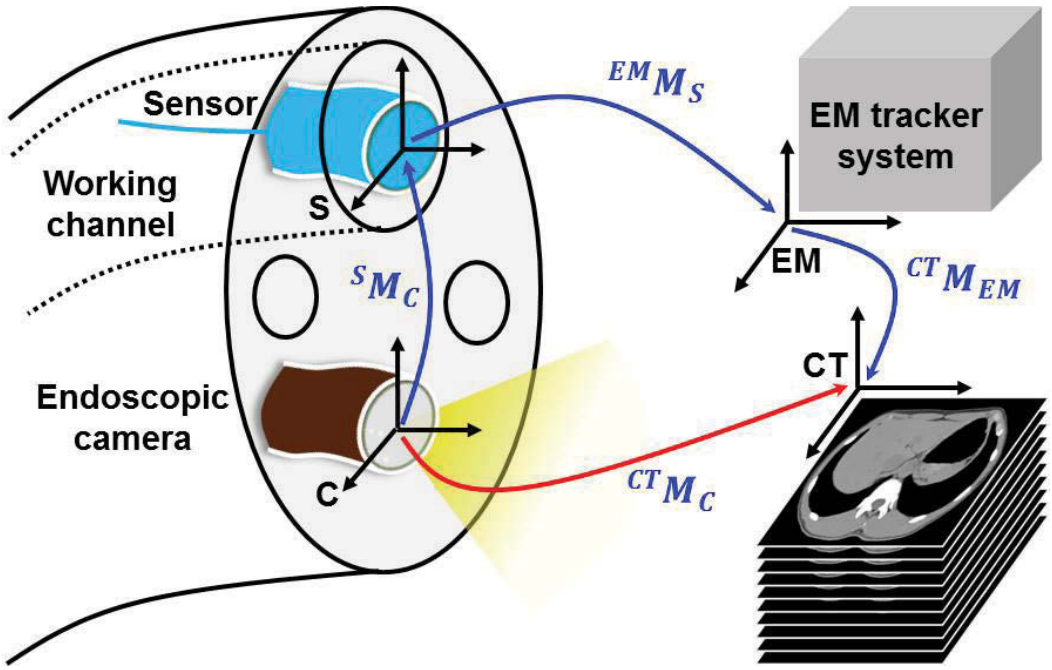


Figure 2.6: Relationships of different coordinate systems in electromagnetically navigated endoscopic procedures

Chapter 3

Calibration and registration

This chapter presents a registration method betweenthe electromagnetic tracker (or sensor) and pre-clinical images which corresponds to our journal paper titled “Adaptive marker-free registration using a multiplepoint strategy for real-time and robust endoscope electromagnetic navigation” that was published in Computer Methods and Programs in Biomedicine. Electromagnetically navigated endoscopic interventions system uses an electromagnetic tracker with a miniature sensor that is usually attached at an endoscope distal tip to real time track endoscope movements in a pre-clinical image space. Our proposed method is an adaptive marker-free registration method that uses a multiple point selection strategy. This method seeks to address an assumption that the endoscope is operated along the centerline of an intraluminal organ which is easily violated during interventions. We introduce an adaptive strategy that generates multiple points in terms of sensor measurements and endoscope tip center calibration. From these generated points, we adaptively choose the optimal point, which is the closest to its assigned the centerline of the hollow organ, to perform registration. A comparative assessment of our new method and the current available methods methods is performed on static and dynamic phantom validation data. Experimental results demonstrate that our proposed adaptive strategy respectively reduced the target registration error from 5.32 to 2.59 mm, and from 7.58 mm to 4.71 mm.

3.1 Coordinate Definitions

The coordinate systems transformation matrix including position and orientation between bronchoscope camera coordinate system to the CT coordinate system need be defined to tracking camera. Figure. 3.1 outlines the relationships and transformation matrices between each coordinate system. ${}^{F}\mathbf{M}_{S}$ describes the relationship between the sensor and magnetic field coordinate systems. ${}^{E}\mathbf{M}_{F}$ is from the magnetic field coordinate system to the EMT

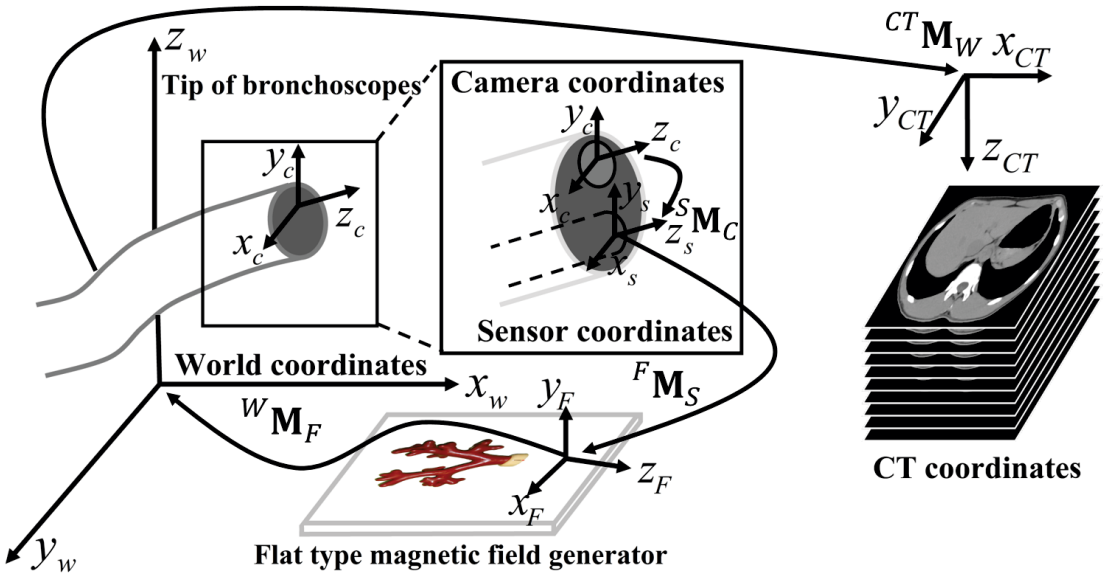


Figure 3.1: Several coordinate systems are defined when using EMT-based methods during bronchoscopic navigation.

coordinate system, and ${}^{CT}\mathbf{M}_E$ is from the EMT coordinate system to the CT coordinate system. We formulate the relationship between the sensor and EMT coordinate systems as ${}^{E}\mathbf{M}_S^{(i)} = {}^{E}\mathbf{M}_F {}^{F}\mathbf{M}_S^{(i)}$, where ${}^{F}\mathbf{M}_S^{(i)}$ is the i-th sensor output. Additionally, the transformation between the camera and the sensor (both attached at the bronchoscope tip) is represented by ${}^{S}\mathbf{M}_C$.

3.1.1 Bronchoscopic Camera Geometry

The mapping between the 3D world (object space) and a 2D image is usually performed by a camera. The pinhole camera model is generally used to characterize the projection from 3D object space onto a 2D image plane. In this model, the camera can be described by an image plane (or focal plane) and a point (the intersection of the principal axis and the image plane) called the central of projection. The minimum distance between the projection center (the origin of the camera coordinate system) and the image plane, which is the focal length of the camera and denoted by f. According to the pinhole model, a world point with coordinates $P_W = (X,Y,Z)^T$ is projected to the point $P_C = (x,y,z)$ in the camera coordinate system: $P_W \mapsto P_C : (X,Y,Z)^T \mapsto (x,y,z)^T = (fX/Z, fY/Z, f)^T$. The pinhole modelling is shown in Figure 3.2.

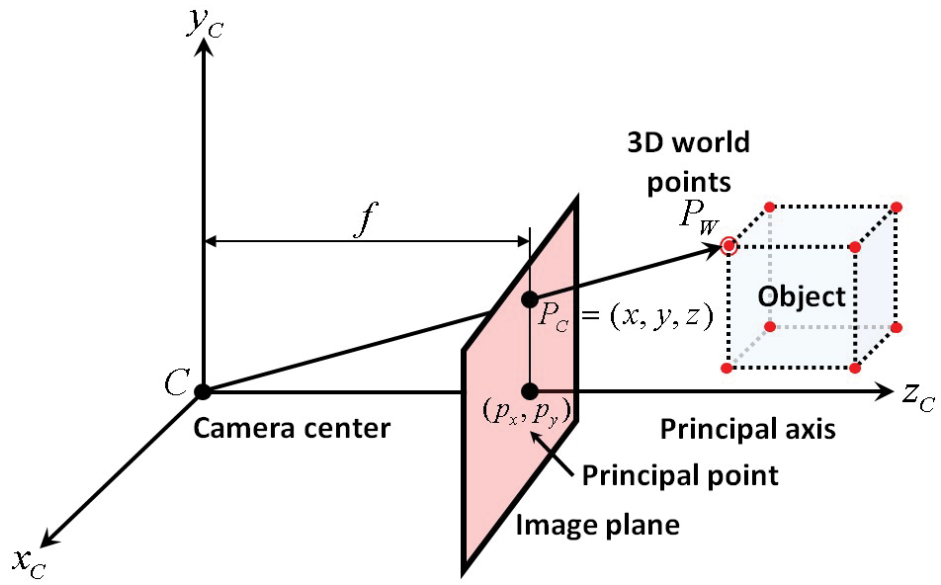


Figure 3.2: Pinhole camera model in the camera coordinate system. C denotes the camera center and is the coordinate origin. The principal point (p_x, p_y) is the intersection of the principal axis and the image plane that is located in front of the focal length.

Let Π be the pinhole mapping function, it projects point P_C in the camera coordinate system onto the image coordinate system as point $p_n = (x_n, y_n, 1)^T$ [37]:

$$\Pi : \Re^3 \mapsto \Re^2, p_n = \lambda \begin{pmatrix} x_n \\ y_n \\ 1 \end{pmatrix} \leftrightarrow P_C = \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} fX + Zp_x \\ fY + Zp_y \\ Z \end{pmatrix} = \mathbf{K} \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} \tag{3.1}$$

where λ is a constant and also means the depth of the world point and $\mathbf{K}$ is defined as the camera intrinsic matrix or parameters and represented by:

$$\mathbf{K} = \begin{pmatrix} f & 0 & p_x \\ 0 & f & p_y \\ 0 & 0 & 1 \end{pmatrix}, \tag{3.2}$$

where (p_x, p_y) is the principal point on the normalized image plane. In practice, due to radial distortion, especially when the camera has a wide field of view, parameters f (including the two focal length of $x-$ and $y-$directions), p_x, and p_y in $\mathbf{K}$ are determined by camera calibration methods, e.g., using Zhang's model [39].

The camera motion including translation and rotation comprises a full six degrees of freedom (DoF) with 3 translational and 3 rotational elements in a 3D space. Any camera pose or motion parameters can be described as some notional point with three direction vectors in

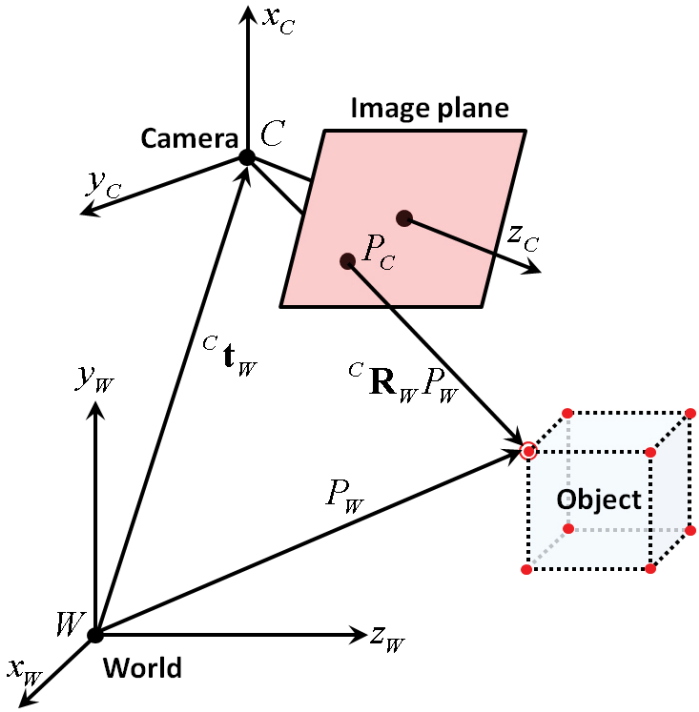


Figure 3.3: Relationship between the world and camera coordinate systems.

the pre-defined coordinate system which is usually called the reference coordinate system (or the world coordinate system) W. The translation part or vector ${}^C\mathbf{t}_W$ depicts the displacement between the reference and the camera coordinate system, usually from the camera center to the origin of the reference coordinate system and is defined as:

$$ {}^C\mathbf{t}_W = \left({}^Ct_W^x, {}^Ct_W^y, {}^Ct_W^z \right)^T, \tag{3.3} $$

where ${}^Ct_W^x$, ${}^Ct_W^y$, and ${}^Ct_W^z$ are the coordinate values of the camera position in the x-, y-, and z-axes of the world coordinate system.

Similarly, the rotation part or matrix ${}^C\mathbf{R}_W$ characterizes the orientation of the camera involved with the reference coordinate frame W. It is constructed by Euler rotation angles θ, ϕ, and ψ of the camera around the x-, y-, and z-axes in the world coordinate system and is defined as:

$$ {}^C\mathbf{R}_W = \begin{pmatrix} C_3C_2 & C_3S_2S_1 - S_3C_1 & C_3S_2C_1 + S_3S_1 \\ S_3C_2 & S_3S_2S_1 + C_3C_1 & S_3S_2C_1 - C_3S_1 \\ -S_2 & C_2S_1 & C_2C_1 \end{pmatrix} \tag{3.4} $$

where the variables of this matrix are computed by: $S_1 = \sin\theta$, $S_2 = \sin\phi$, $S_3 = \sin\psi$, $C_1 = \cos\theta$, $C_2 = \cos\phi$, and $C_3 = \cos\psi$. By the pinhole model, points may be transformed

from the reference coordinate system to the camera coordinate system. This is represented in Figure 3.3.

From the vector triangle in Figure 3.3: ${}^{C}\mathbf{R}_{W}P_{C}+{}^{C}\mathbf{t}_{W}-P_{W}=0$, the camera extrinsic pose $({}^{W}\mathbf{R}_{C}\mid{}^{W}\mathbf{t}_{C})$ is proved to satisfy the following formulation [37]:

$$P_{C}=({}^{C}\mathbf{R}_{W})^{-1}(P_{W}-{}^{C}\mathbf{t}_{W})={}^{W}\mathbf{R}_{C}P_{W}-({}^{C}\mathbf{R}_{W})^{-1}({}^{C}\mathbf{t}_{W})={}^{W}\mathbf{R}_{C}P_{W}+{}^{W}\mathbf{t}_{C}, \tag{3.5}$$

that is, the motion parameters $({}^{W}\mathbf{R}_{C}\mid{}^{W}\mathbf{t}_{C})$ is used to register two coordinate systems from the camera to the reference coordinate systems.

By combining Eq. 3.2, the camera pose or matrix $\mathbf{M}$ including the intrinsic and extrinsic parameters can be represented by:

$$\mathbf{M}=\mathbf{K}\left[{}^{W}\mathbf{R}_{C}\ \mid\ {}^{W}\mathbf{t}_{C}\right], \tag{3.6}$$

that is, $\mathbf{M}$ maps the image and the reference (or world) coordinate systems. The goal of this thesis is to continuously estimate or calculate the camera pose $\mathbf{M}$ between pre- and intra-operative imaging coordinate systems.

3.2 Hand-eye calibration

Utilizing hand-eye calibration was introduced for a variety of robotic applications.The cameras was calibrated using a calibration pattern. Then the unknown transformation from the robot coordinate system to the calibration pattern coordinate system as well as the transformation from the camera to the hand coordinate system need to be estimated simultaneously.

3.2.1 Tsai's method

We must calculate the rigid transformation matrix ${}^{S}\mathbf{M}_{C}$ between the bronchoscope camera and the EM sensor coordinate systems. For that, we utilize hand-eye calibration. It computes the spatial relationship from the hand (e.g. a robot gripper) to the eye (e.g. a camera) that is an unknown rigid transformation which can be determined from a number of acquired camera and gripper motions. In our situation, the "*hand*" corresponds to the EM sensor while the "*eye*" relates to the bronchoscope camera. We here utilize the hand-eye calibration method to calculate the transformation between the EM sensor coordinate system and the

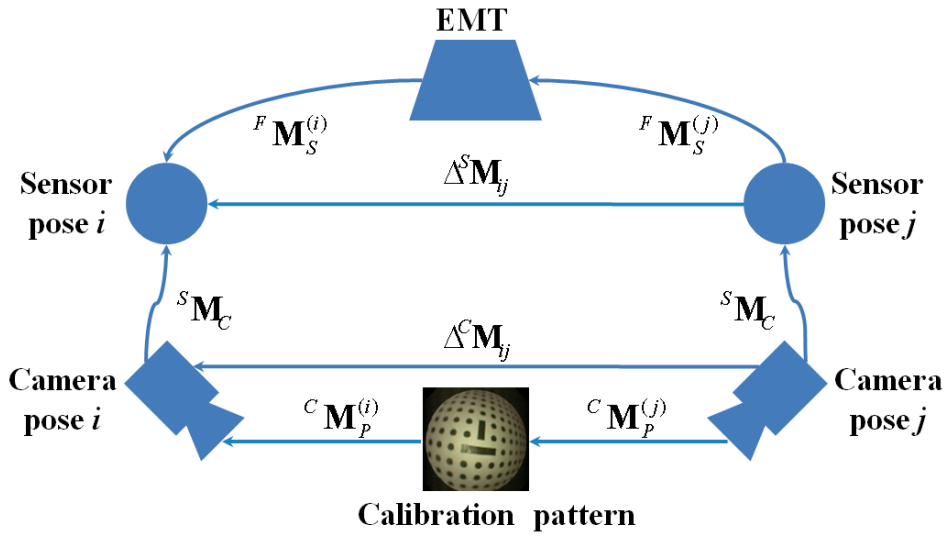


Figure 3.4: Hand-eye calibration setup from camera pose estimation and EM sensor outputs.

camera coordinate system by

$$\Delta^{S}\mathbf{M}_{ij}\,{}^{S}\mathbf{M}_{C} = {}^{S}\mathbf{M}_{C}\Delta^{C}\mathbf{M}_{ij} \tag{3.7}$$

where $\Delta^{S}\mathbf{M}_{ij}$ is the sensor movement relative to the EM tracker coordinate system $\Delta^{S}\mathbf{M}_{ij} = {}^{S}\mathbf{M}_{F}^{(i)}\,{}^{F}\mathbf{M}_{S}^{(j)}$, and $\Delta^{C}\mathbf{M}_{ij}$ is the camera motion relative to a calibration pattern frame $\Delta^{C}\mathbf{M}_{ij} = {}^{C}\mathbf{M}_{P}^{(i)}\,{}^{P}\mathbf{M}_{C}^{(j)}$, for the bronchoscope performing a movement from the i-th pose to the j-th pose. Equation 3.7 can be directly obtained from Figure 3.4.

During hand-eye calibration we first acquire pairs of calibration pattern images and EM sensor outputs. We next perform a calibration of the intrinsic and distortion parameters of the bronchoscope camera and estimate the camera pose $^{C}\mathbf{M}_{P}^{(i)}$ for each frame i utilizing Zhang's method[39]. Then, we solve Equation 3.7 to obtain the transformation $^{S}\mathbf{M}_{C}$ [40].

After hand-eye calibration, we calculate an initial transformation from the world (patient) coordinate system to the CT coordinate system by localizing corresponding points in the EM tracking and CT coordinate system and computing the rigid transformation between the two point clouds [41] as outlined in the next section.

3.3 EM-CT registration

electromagnetically navigated endoscopy, which uses an ET system to track endoscope movements, is employed to address such a challenge in operating rooms, e.g., a commercially available superDimension navigation system(superDimension Inc., USA), in which the

diagnostic accuracy has been reported about 59–74% without sensitivity to tumor size [32, 33]. During the electromagnetic endoscope navigation, an ET position sensor is usually attached at the endoscope distal tip to measure endoscope movements. To transform these ET sensor measurements to the CT volume,the spatial alignment between the ET system and CT volume must be determined

Currently several methods have been discussed to compute the ET-CT transformation in the literature. Schwarz et al. [4] introduced a marker-based registration. They manually chose among five and seven anatomic markers at the bronchial bifurcations to establish the ET-CT alignment and reported the registration accuracy about 5.7 mm under a patient movement compensation mechanism that utilized three ET sensors attached to patient thoraxes. Without using anatomical markers, Klein et al. [2] maximized the percentage of ET sensor measurements inside the CT airway volume to determine the ET-CT transformation. Unfortunately, such a method depends heavily on its initial estimate of the ET-CT alignment in maximization. It possibly fails to converge to an optimal solution during optimization. Based on the centerlines of the airway trees, Deguchi et al. [1] proposed a marker-free method that minimizes the distance between the sensor measurements and bronchial centerlines to obtain the ET-CT transformation. They assumed that the endoscope is operated along the centerlines of the bronchi. However, surgeons may easily break such an hypothesis in endoscopic examinations or interventions.

In general, marker-free registration approaches are promising for endoscope electromagnetic navigation since they are easily performed in operating rooms without manually choosing any anatomical markers in the body. The motivation of this work is to diminish the assumption of moving endoscopes along the centerlines of hollow organs (e.g., airway trees and nose) in the body, since such an assumption implies that the closer to the centerlines of operating the endoscope, the more accurate ET-CT transformation can be obtained. Our basic idea is to generate multiple positions or points around the endoscope shaft on the basis of each electromagnetic sensor measurement (position and orientation); points closest to the centerlines are selected for registration. We call such a procedure as a multiple point strategy that was proved to be very effective to estimate the ET-CT alignment. The preliminary work of this paper was orally presented at a conference [42].

3.3.1 Adaptive marker-free registration

This section describes our adaptive marker-free registration method. Basically, it is comprised of several steps: (1) CT image segmentation and initialization of optimization, (2) endoscope center calibration, (3) multiple point strategy for selecting optimal points, and (4) optimization for determining the ET-CT transformation. Each step is detailed as follows.

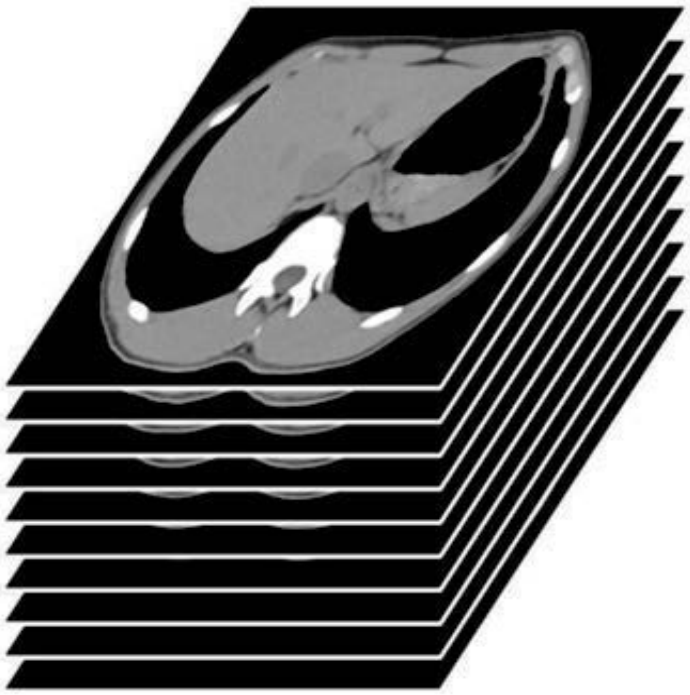


Figure 3.5: CT images for the airway tree voume segmentation

Segmentation and Initialization

Due to the hollow organ's centerlines that are used in our method, we first segment CT images to obtain these centerlines (Figs. 3.5 $\sim$ 3.7) that are described by $\mathscr{C} = \{\mathbf{c}_k = (\mathbf{c}_k^s, \mathbf{c}_k^e)\}_{k=1}^K$, where each centerline $\mathbf{c}_k$ has its start and end points, $\mathbf{c}_k^s$ and $\mathbf{c}_k^e$, and K is the number of extracted centerlines [29].

We must initialize ET-CT transformation ${}^{ct}\mathscr{T}_{et}$ before estimating it during optimization. We use three centerlines to perform such an initialization. The carina position $\mathbf{q}_{ct}$, which is easily selected from CT images, is located around the end point of the trachea centerline $\mathbf{c}_t$ or the start points of the left and right primary bronchus centerlines $\mathbf{c}_l$ and $\mathbf{c}_r$: $\mathbf{q}_{ct} = \mathbf{c}_t^e = \mathbf{c}_l^s = \mathbf{c}_r^s$. We locate the endoscope with an attached ET sensor at the carina of the trachea during navigation and obtain sensor measurement $\mathbf{m}_{et} = (\mathbf{p}_{et}, \mathbf{r}_{et})$, where $\mathbf{p}_{et}$ is the sensor position and $\mathbf{r}_{et}$ denotes the sensor orientation with $\mathbf{s}_{et}^x$, $\mathbf{s}_{et}^y$, and $\mathbf{s}_{et}^z$ in the three directions of x-, y-, and z-axes: $\mathbf{r}_{et} = (\mathbf{s}_{et}^x, \mathbf{s}_{et}^y, \mathbf{s}_{et}^z)$. Hence initial transformation ${}^{ct}\mathscr{T}_{et}^0$ before optimization can be calculated by:

$$ {}^{ct}\mathscr{T}_{et}^0 = \begin{pmatrix} \mathbf{r}_{ct} & \mathbf{q}_{ct} \\ \mathbf{0}^T & 1 \end{pmatrix} \begin{pmatrix} \mathbf{r}_{et} & \mathbf{p}_{et} \\ \mathbf{0}^T & 1 \end{pmatrix}^{-1}, \tag{3.8} $$

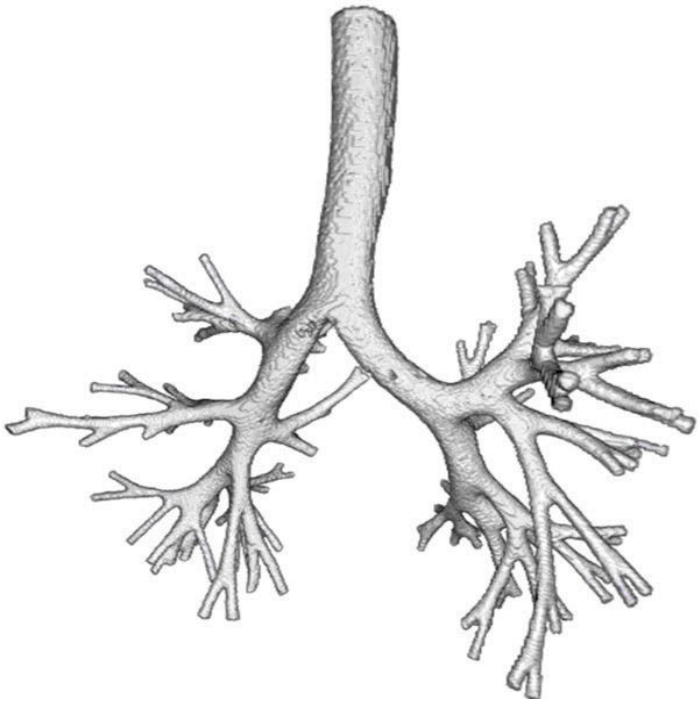


Figure 3.6: Automatically segmented airway volume from CT images

where orientation $\mathbf{r}_{ct}$ includes three directions of $\mathbf{s}_{ct}^{x}$, $\mathbf{s}_{ct}^{y}$, and $\mathbf{s}_{ct}^{z}$ and these directional vectors are computed by:

$$\mathbf{s}_{ct}^{z} = \frac{\mathbf{c}_t^e - \mathbf{c}_t^s}{||\mathbf{c}_t^e - \mathbf{c}_t^s||},\ \mathbf{s}_{ct}^{y} = \frac{\mathbf{c}_l^e - \mathbf{c}_l^s}{||\mathbf{c}_l^e - \mathbf{c}_l^s||} \times \frac{\mathbf{c}_r^e - \mathbf{c}_r^s}{||\mathbf{c}_r^e - \mathbf{c}_r^s||}, \tag{3.9}$$

and $\mathbf{s}_{ct}^{x} = \mathbf{s}_{ct}^{y} \times \mathbf{s}_{ct}^{z}$, where symbol $\times$ denotes the cross product. Equation (2) indicates that our initialization involves the rough orientation information. Our proposed algorithm on the basis of such an initialization of the ET-CT transformation usually succeeds to converge. The speed of convergence depends on the accuracy of the initialization.

Endoscope Center Calibration

To generate multiple points, we must determine the relationship between the ET sensor and the center of the endoscope tip transversal plane. To calibrate this relationship, we build a model and insert the endoscope tip with an integrated ET sensor into this model and spin the rotatable cuboid (Figs. 3.8 $\sim$ 3.10). Note that the endoscope is rotated along its centers and endoscopes with different diameters require different sizes of calibration models.

We acquire three measurements of the ET sensor after spinning the rotatable cuboid every 120° and represent these measurements by: $\mathbf{m}_a = (\mathbf{p}_a, \mathbf{s}_a^x, \mathbf{s}_a^y, \mathbf{s}_a^z)$, $\mathbf{m}_b = (\mathbf{p}_b, \mathbf{s}_b^x, \mathbf{s}_b^y, \mathbf{s}_b^z)$, and $\mathbf{m}_c = (\mathbf{p}_c, \mathbf{s}_c^x, \mathbf{s}_c^y, \mathbf{s}_c^z)$. Actually, we moved the rotatable cuboid and acquired about 100

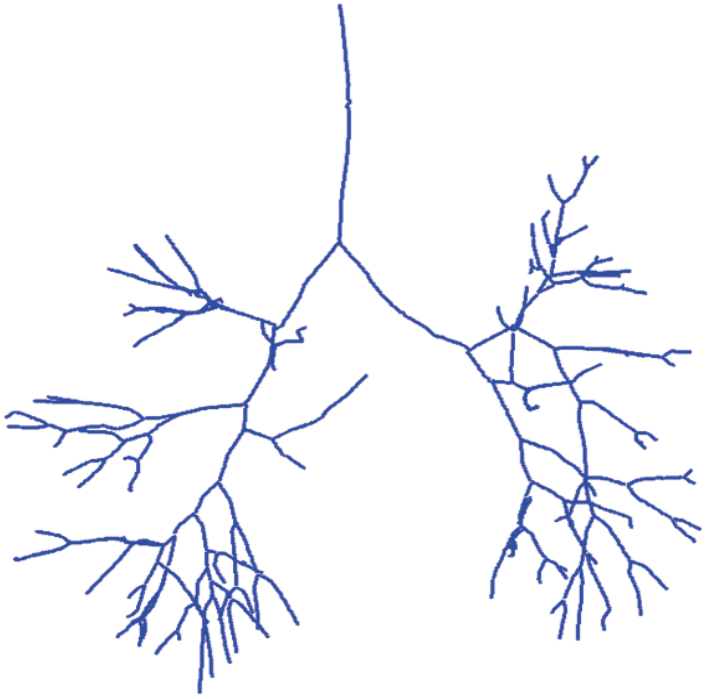


Figure 3.7: Extracted airway centerlines from the segmented airway

sensor outputs at each position and determined positions $\mathbf{p}_a$, $\mathbf{p}_b$, and $\mathbf{p}_c$ by the average of these sensor outputs. Since these average positions, $\mathbf{p}_a = (p_a^x, p_a^y, p_a^z)^T$, $\mathbf{p}_b = (p_b^x, p_b^y, p_b^z)^T$, and $\mathbf{p}_c = (p_c^x, p_c^y, p_c^z)^T$, should be on one circle plane centered at point $\mathbf{p}_o$, we have:

$$||\mathbf{p}_o - \mathbf{p}_a|| = ||\mathbf{p}_o - \mathbf{p}_b|| = ||\mathbf{p}_o - \mathbf{p}_c||, \tag{3.10}$$

by solving this equation, we obtain center point $\mathbf{p}_o$. Since the spatial relationship between the ET sensor and the endoscope tip center was relatively fixed, it can move point $\mathbf{p}_a$ with typical spatial distance $(\delta_x, \delta_y, \delta_z)$ along the ET sensor orientation $(\mathbf{s}_a^x, \mathbf{s}_a^y, \mathbf{s}_a^z)$ to the endoscope tip center $\mathbf{p}_o = (p_o^x, p_o^y, p_o^z)^T$:

$$\begin{pmatrix} \overbrace{p_o^x - p_a^x}^{\mathbf{p}_o - \mathbf{p}_a} \\ p_o^y - p_a^y \\ p_o^z - p_a^z \end{pmatrix} = \begin{pmatrix} \overbrace{s_a^{x,1}}^{\mathbf{s}_a^x} & \overbrace{s_a^{y,1}}^{\mathbf{s}_a^y} & \overbrace{s_a^{z,1}}^{\mathbf{s}_a^z} \\ s_a^{x,2} & s_a^{y,2} & s_a^{z,2} \\ s_a^{x,3} & s_a^{y,3} & s_a^{z,3} \end{pmatrix} \begin{pmatrix} \delta_x \\ \delta_y \\ \delta_z \end{pmatrix}. \tag{3.11}$$

By solving Eq. 3.11, we determine spatial relationship $(\delta_x, \delta_y, \delta_z)$ between the ET sensor and the endoscope tip center.

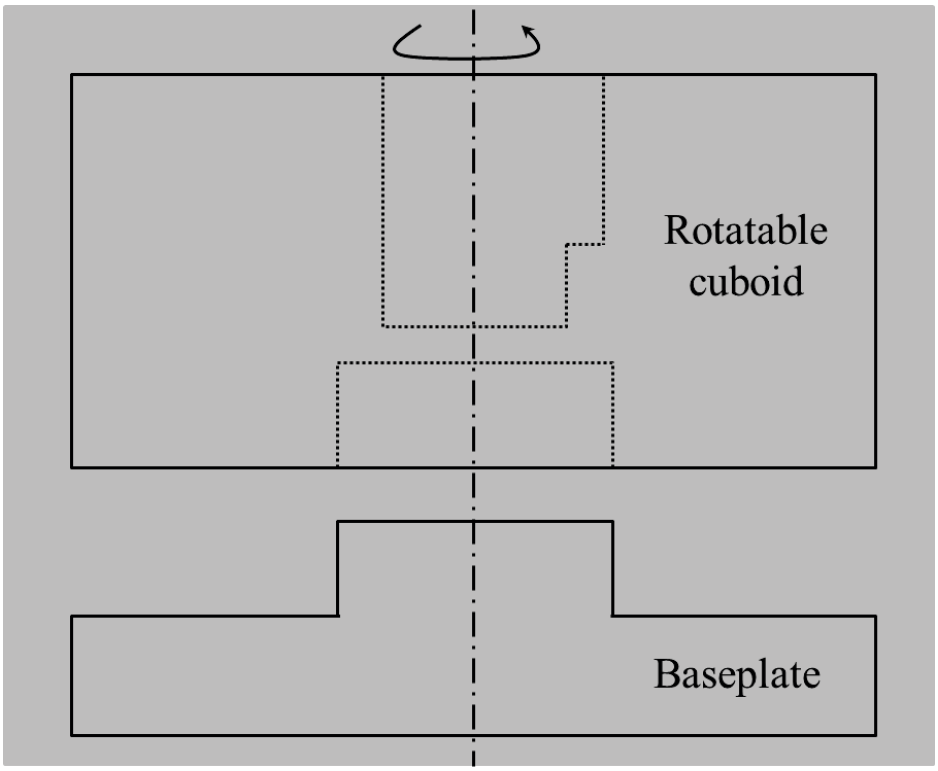


Figure 3.8: The model drawing for calibrating the spatial relationship between the ET sensor and the endoscope tip center

Based on spatial distance (δ_x, δ_y, δ_z) and the sensor orientation, endoscope tip center position $\hat{\mathbf{p}}_i$, which corresponds to measurement $\mathbf{m}_i = (\mathbf{p}_i, \mathbf{s}_i^x, \mathbf{s}_i^y, \mathbf{s}_i^z)$ at time i, is computed by:

$$\hat{\mathbf{p}}_i = \mathbf{p}_i + \delta_x \underbrace{\begin{pmatrix} s_i^{x,1} \\ s_i^{x,2} \\ s_i^{x,3} \end{pmatrix}}_{\mathbf{s}_i^x} + \delta_y \underbrace{\begin{pmatrix} s_i^{y,1} \\ s_i^{y,2} \\ s_i^{y,3} \end{pmatrix}}_{\mathbf{s}_i^y} + \delta_z \underbrace{\begin{pmatrix} s_i^{z,1} \\ s_i^{z,2} \\ s_i^{z,3} \end{pmatrix}}_{\mathbf{s}_i^z}. \tag{3.12}$$

Using calibrated endoscope tip center position $\hat{\mathbf{p}}_i$ and sensor measured point or position $\mathbf{p}_i$, we generate multiple points for marker-free registration, as discussed as follows.

Multiple Point Strategy

After obtaining $\hat{\mathbf{p}}_i$ and current measured point $\mathbf{p}_i$, we generate multiple points. Our basic idea is to divide the endoscope tip circumference into eight equal parts and generate candidate points $\{\mathbf{p}_i^1, \mathbf{p}_i^2, \mathbf{p}_i^3, \mathbf{p}_i^4, \mathbf{p}_i^5, \mathbf{p}_i^6, \mathbf{p}_i^7\}$, as shown in Figs. 3.11 $\sim$ 3.13.

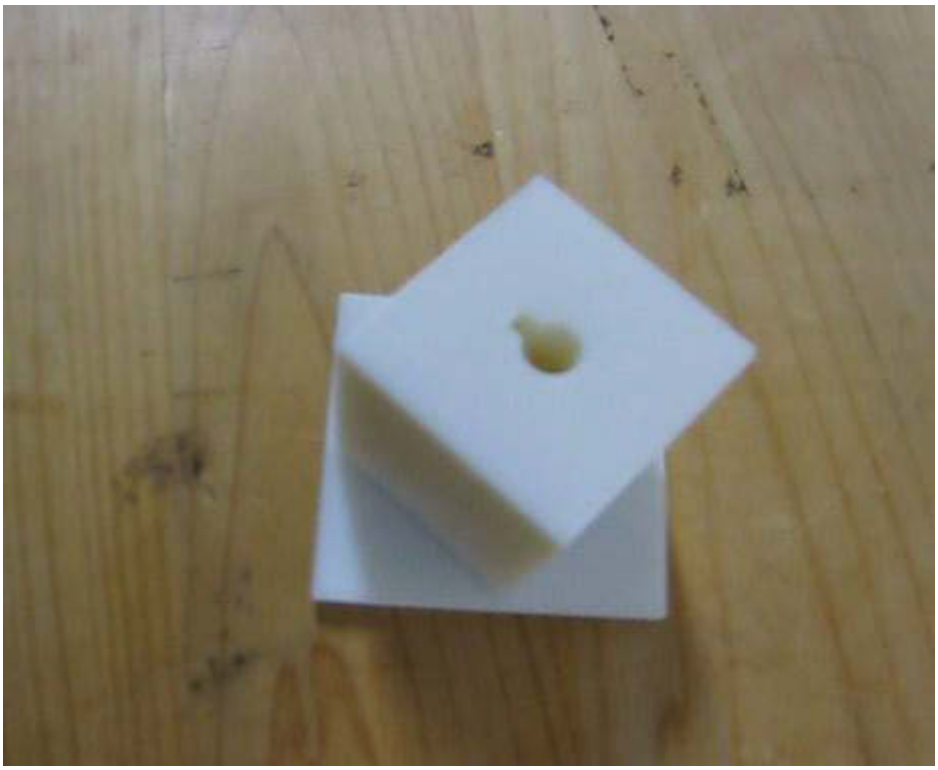


Figure 3.9: The calibration model was produced by using a 3-D printer

Based on Euclidean geometry, we first compute the position of points $\mathbf{p}_i^4$, $\mathbf{p}_i^2$, and $\mathbf{p}_i^6$ by ($r = ||\hat{\mathbf{p}}_i - \mathbf{p}_i||$):

$$\begin{cases} \mathbf{p}_i^4 = \hat{\mathbf{p}}_i + (\hat{\mathbf{p}}_i - \mathbf{p}_i) \\ \mathbf{p}_i^2 = \hat{\mathbf{p}}_i + r\left(\frac{\hat{\mathbf{p}}_i - \mathbf{p}_i}{\|\hat{\mathbf{p}}_i - \mathbf{p}_i\|} \times \mathbf{s}_i^z\right) \\ \mathbf{p}_i^6 = \hat{\mathbf{p}}_i - r\left(\frac{\hat{\mathbf{p}}_i - \mathbf{p}_i}{\|\hat{\mathbf{p}}_i - \mathbf{p}_i\|} \times \mathbf{s}_i^z\right) \end{cases} . \tag{3.13}$$

Therefore, the position of the other four points, $\mathbf{p}_i^1$, $\mathbf{p}_i^3$, $\mathbf{p}_i^5$, and $\mathbf{p}_i^7$ is determined by (symbol $\times$ means the cross product):

$$\begin{cases} \mathbf{p}_i^1 = \hat{\mathbf{p}}_i - r\left(\frac{\mathbf{p}_i^2 - \mathbf{p}_i}{\|\mathbf{p}_i^2 - \mathbf{p}_i\|} \times \mathbf{s}_i^z\right) \\ \mathbf{p}_i^3 = \hat{\mathbf{p}}_i - r\left(\frac{\mathbf{p}_i^6 - \mathbf{p}_i}{\|\mathbf{p}_i^6 - \mathbf{p}_i\|} \times \mathbf{s}_i^z\right) \\ \mathbf{p}_i^5 = \hat{\mathbf{p}}_i + r\left(\frac{\mathbf{p}_i^2 - \mathbf{p}_i}{\|\mathbf{p}_i^2 - \mathbf{p}_i\|} \times \mathbf{s}_i^z\right) \\ \mathbf{p}_i^7 = \hat{\mathbf{p}}_i + r\left(\frac{\mathbf{p}_i^6 - \mathbf{p}_i}{\|\mathbf{p}_i^6 - \mathbf{p}_i\|} \times \mathbf{s}_i^z\right) \end{cases} . \tag{3.14}$$

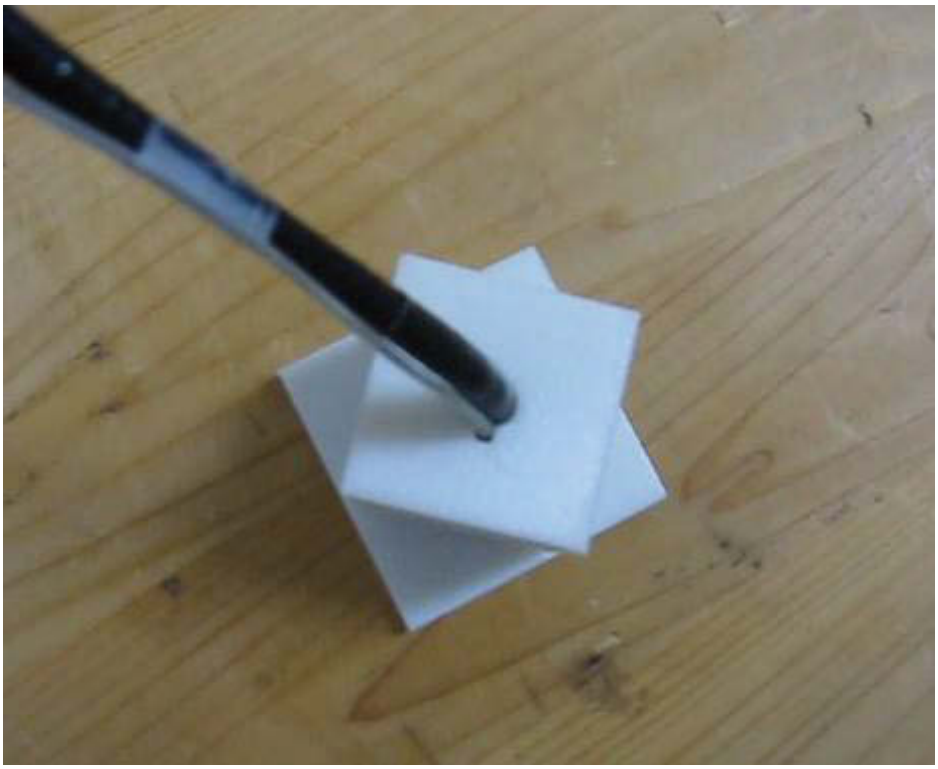


Figure 3.10: Insert the endoscope with an integrated ET sensor into the calibration model

To solve Equations (6) and (7), we assume that the ET sensor and endoscope tip center are in the same plane of rotation.

Finally, nine points $\Omega = \{\mathbf{p}_i, \hat{\mathbf{p}}_i, \mathbf{p}_i^1, \mathbf{p}_i^2, \mathbf{p}_i^3, \mathbf{p}_i^4, \mathbf{p}_i^5, \mathbf{p}_i^6, \mathbf{p}_i^7\}$ are obtained in the ET coordinate system. We select the optimal one from set Ω to compute spatial alignment ${}^{ct}\mathscr{T}_{et}$.

We transform ET point $\mathbf{p}_i$ to CT point ${}^{ct}\mathscr{T}_{et}\mathbf{p}_i$ and assign closest centerline $\hat{\mathbf{c}}_k$ to ${}^{ct}\mathscr{T}_{et}\mathbf{p}_i$ by minimizing the distance between point ${}^{ct}\mathscr{T}_{et}\mathbf{p}_i$ and centerlines $\mathscr{C}$.

$$\hat{\mathbf{c}}_k = \arg\min_{\mathbf{c}_k \in \mathscr{C}} d({}^{ct}\mathscr{T}_{et}\mathbf{p}_i, \mathbf{c}_k), \tag{3.15}$$

where distance $d({}^{ct}\mathscr{T}_{et}\mathbf{p}_i, \mathbf{c}_k)$ from point ${}^{ct}\mathscr{T}_{et}\mathbf{p}_i$ to centerline $\mathbf{c}_k$ is calculate by (Fig. 3.13):

$$d({}^{ct}\mathscr{T}_{et}\mathbf{p}_i, \mathbf{c}_k) = \begin{cases} ||{}^{ct}\mathscr{T}_{et}\mathbf{p}_i - \mathbf{c}_k^s|| & \gamma < 0 \\ ||{}^{ct}\mathscr{T}_{et}\mathbf{p}_i - \mathbf{c}_k^e|| & \gamma > l_k \\ \sqrt{||{}^{ct}\mathscr{T}_{et}\mathbf{p}_i - \mathbf{c}_k^s||^2 - \gamma^2} & \text{otherwise} \end{cases}, \tag{3.16}$$

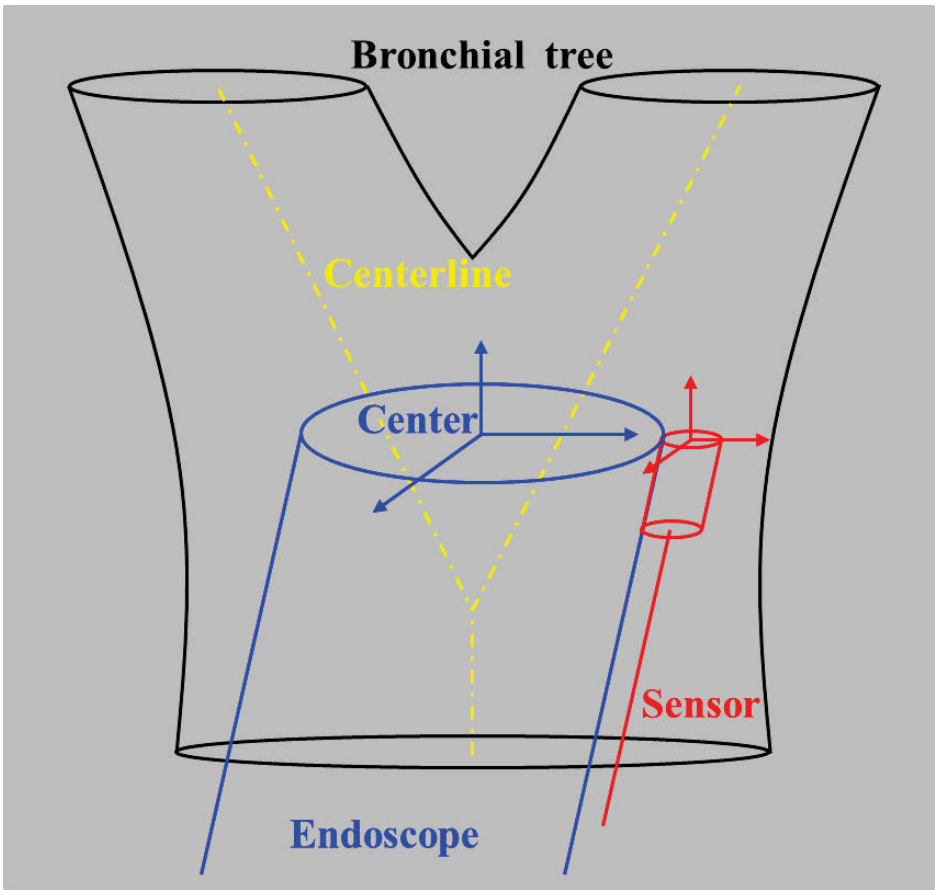


Figure 3.11: Attached an ET sensor at the endoscope distal tip and calibrate the endoscope tip center

where the centerline length $l_k = ||\mathbf{c}_k^e - \mathbf{c}_k^s||$ and γ denotes the length of vector $({}^{ct}\mathscr{T}_{et}\mathbf{p}_i - \mathbf{c}_k^s)$ that is projected on centerline $\mathbf{c}_k$ and is computed by:

$$\gamma = \frac{\left\langle {}^{ct}\mathscr{T}_{et}\mathbf{p}_i - \mathbf{c}_k^s,\ \mathbf{c}_k^e - \mathbf{c}_k^s \right\rangle}{||\mathbf{c}_k^e - \mathbf{c}_k^s||}, \tag{3.17}$$

where $\gamma < 0$ and $\gamma > l_k$ mean that projected point $\mathbf{q}_i$ is on the parent and child centerlines of centerline $\mathbf{c}_k$, respectively; otherwise, it is on centerline $\mathbf{c}_k$, and $\langle , \rangle$ denotes the dot product.

Note that during the minimization (Eq. 3.15), we may obtain several closest centerlines $\{\hat{\mathbf{c}}_{k,j}\}_{j=1,2,3,\ldots}$ that have the same closest distance to CT point ${}^{ct}\mathscr{T}_{et}\mathbf{p}_i$. We compute the angle between sensor $z-$direction $\mathbf{s}_i^z$ and centerline direction $(\hat{\mathbf{c}}_{k,j}^e - \hat{\mathbf{c}}_{k,j}^s)||\hat{\mathbf{c}}_{k,j}^e - \hat{\mathbf{c}}_{k,j}^s||^{-1}$ to determine optimal centerline $\tilde{\mathbf{c}}_j$:

$$\tilde{\mathbf{c}}_j = \arg\min_{\{\hat{\mathbf{c}}_{k,j}\}} \arccos \left\langle \frac{\hat{\mathbf{c}}_{k,j}^e - \hat{\mathbf{c}}_{k,j}^s}{||\hat{\mathbf{c}}_{k,j}^e - \hat{\mathbf{c}}_{k,j}^s||}, \frac{{}^{ct}\mathscr{T}_{et}\mathbf{s}_i^z}{||{}^{ct}\mathscr{T}_{et}\mathbf{s}_i^z||} \right\rangle. \tag{3.18}$$

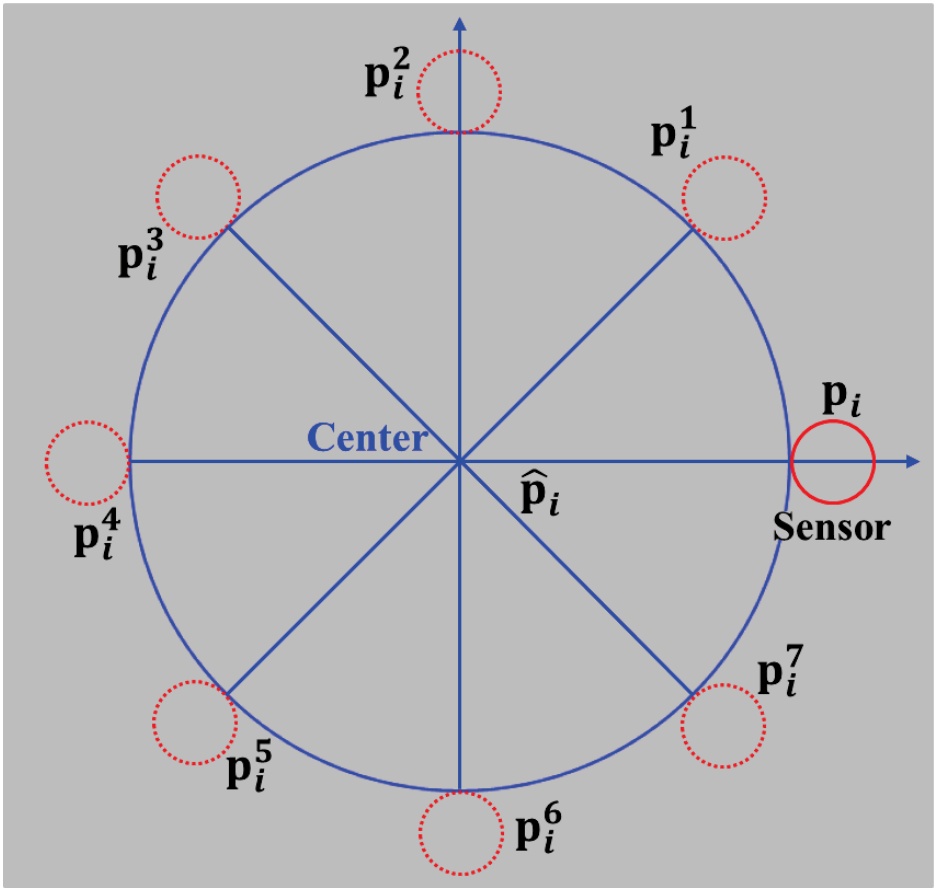


Figure 3.12: Generated multiple points on basis of sensor measurement $\mathbf{p}_i$ and endoscope tip center $\hat{\mathbf{p}}_i$, and assign centerline $\mathbf{c}_k$ to transformed point ${}^{ct}\mathscr{T}_{et}\mathbf{p}_i$

After obtaining optimal centerline $\tilde{\mathbf{c}}_k$, we select optimal point $\tilde{\mathbf{p}}_i$ from set Ω in terms of distance $d({}^{ct}\mathscr{T}_{et}\hat{\mathbf{p}}_i, \hat{\mathbf{c}}_k)$:

$$\tilde{\mathbf{p}}_i = \arg\min_{\hat{\mathbf{p}}_i \in \Omega} d({}^{ct}\mathscr{T}_{et}\hat{\mathbf{p}}_i, \tilde{\mathbf{c}}_j). \tag{3.19}$$

Note that $\tilde{\mathbf{p}}_i$ is the output of our multiple point strategy.

Determination of Transformation

After determining optimal centerline $\tilde{\mathbf{c}}_j$ and point $\tilde{\mathbf{p}}_i$ at time i, we project point $\tilde{\mathbf{p}}_i$ on centerline $\tilde{\mathbf{c}}_j$ and compute projected point $\mathbf{q}_i$ by (Fig. 3.13):

$$\mathbf{q}_i = \tilde{\mathbf{c}}_j^s + \frac{\left\langle {}^{ct}\mathscr{T}_{et}\tilde{\mathbf{p}}_i - \tilde{\mathbf{c}}_j^s,\ \tilde{\mathbf{c}}_j^e - \tilde{\mathbf{c}}_j^s \right\rangle}{||\tilde{\mathbf{c}}_j^e - \tilde{\mathbf{c}}_j^s||} \frac{(\tilde{\mathbf{c}}_j^e - \tilde{\mathbf{c}}_j^s)}{||\tilde{\mathbf{c}}_j^e - \tilde{\mathbf{c}}_j^s||}. \tag{3.20}$$

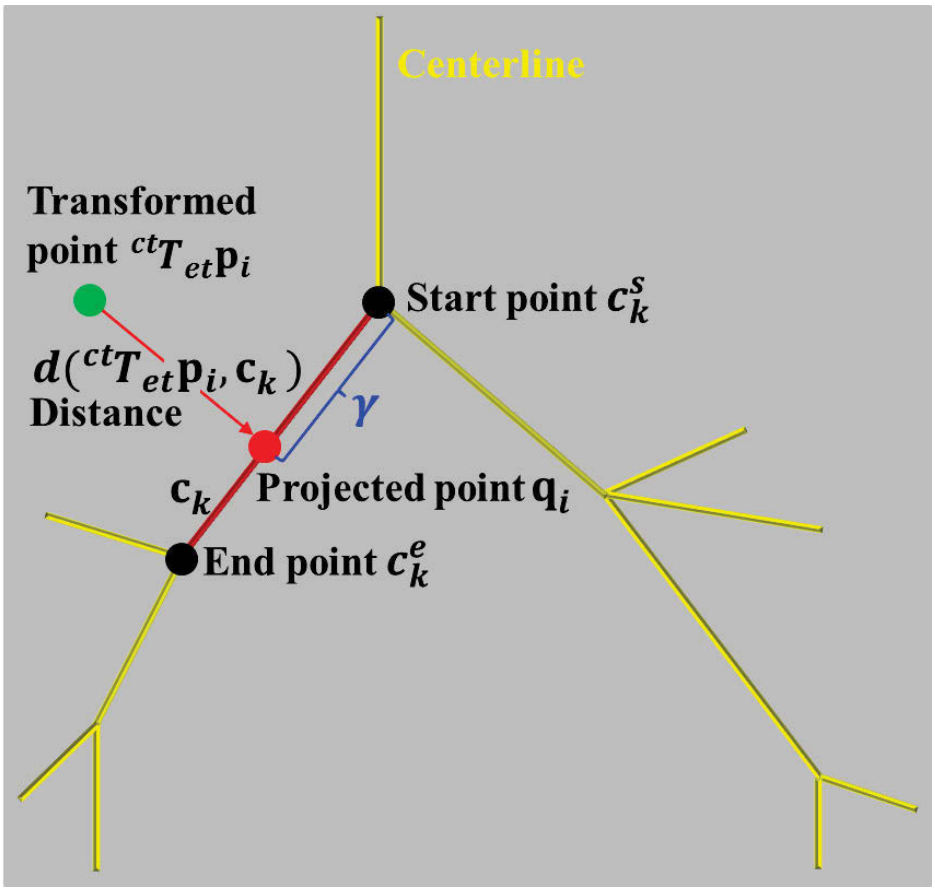


Figure 3.13: Obtained projected point $\mathbf{q}_i$ in terms of distance $d({}^{ct}\mathscr{T}_{et}\mathbf{p}_i, \mathbf{c}_k)$

For M sensor measurements, we use Eqs. 3.13∼3.20 and obtain point set $\mathscr{M} = \{\tilde{\mathbf{m}}_i = (\tilde{\mathbf{p}}_i, \mathbf{s}_i^x, \mathbf{s}_i^y, \mathbf{s}_i^z)\}_{i=1}^M$ and projected point set $\mathscr{Q} = \{\mathbf{q}_i\}_{i=1}^M$.

Based on two point sets $\mathscr{M}$ and $\mathscr{Q}$, we run an optimizer to estimate spatial transformation ${}^{ct}\mathscr{T}_{et}$. However, an objective function must be defined in this optimization. Since an endoscope is assumed to manipulate along the centerlines, it is natural to define objective function $\mathscr{D}_i$ with respect to the distance between transformed point ${}^{ct}\mathscr{T}_{et}\tilde{\mathbf{p}}_i$ and projected point $\mathbf{q}_i$ and radius ϕ_j of optimal centerline $\tilde{\mathbf{c}}_j$:

$$\mathscr{D}_i = \frac{||{}^{ct}\mathscr{T}_{et}\tilde{\mathbf{p}}_i - \mathbf{q}_i||}{\phi_j}. \tag{3.21}$$

This equation also implies that the point closer to the centerline, the more accurate estimation ${}^{ct}\mathscr{T}_{et}$ can be obtained.

Eventually, ${}^{ct}\mathscr{T}_{et}$ is determined by minimizing the distance between selected point set $\mathscr{M} = \{\tilde{\mathbf{m}}_i = (\tilde{\mathbf{p}}_i, \mathbf{s}_i^x, \mathbf{s}_i^y, \mathbf{s}_i^z)\}_{i=1}^M$ and projected point set $\mathscr{Q} = \{\mathbf{q}_i\}_{i=1}^M$ (M is the number of total

ET sensor measurements):

$$^{ct}\tilde{\mathscr{T}}_{et} = \arg \min_{\tilde{\mathbf{m}}_i \in \mathscr{M}, \mathbf{q}_i \in \mathscr{Q}} \sum_i \mathscr{D}_i. \tag{3.22}$$

The above equation that formulates a minimization procedure was solved by the Powell approach [43] in our implementation.

3.3.2 Experiments

We evaluate our adaptive marker-free registration approach on static and dynamic phantoms and use different types of endoscopes and ET systems in our experiments.

We built a static phantom that can not simulate respiratory motion to validate our proposed registration method. The CT volume of this static phantom was $512 \times 512 \times 611$ voxels with resolution of $0.892 \times 0.692 \times 0.5$ mm^3. During this static phantom validation, we use an endobronchial ultrasound (EBUS) endoscope (BF TYPE US260F-OL8, Olympus, Tokyo) and an ET system (AURORA, Northern Digital Inc., Waterloo, Canada) (Fig. 3.14). The radius of this EBUS endoscope and an AURORA sensor was about 3.5 mm and 0.4 mm. We attached the AURORA sensor at the EBUS endoscope tip.

We also experimentally constructed a dynamic phantom that can simulate respiratory motion with a breath rate range of 0∼10 per minute and a motion domain of 0∼24 mm. The phantom CT acquisition parameters were 512×512 pixels, 1021 slices, a 0.68 mm reconstruction pitch, and 0.5 mm thick slices. One drawback of our dynamic phantom must be noted; its breath rates are somewhat unrealistic, compared to human breath rates of 14∼30. Our previous work showed more details about constructing our dynamic phantom [6]. During this dynamic phantom validation, we used a 3-D Guidance medSAFE tracker (Ascension Technology Corporation, USA) as our ET system with an electromagnetic field generator and sensors (1.5 mm, 6DoF) and an endoscope (BF Type 200, Olympus, Tokyo) (Fig. 3.15). The tip diameter of this endoscope was 5.7 mm.

For a lateral comparison, we investigate several currently available methods: (1) Kein et al. [2], an in-volume maximization algorithm that maximizes the percentage of ET sensor measurements inside the airway volume, (2) Deguchi et al. [1], an approach that minimizes the distance between ET sensor measurements and centerlines of the airway trees, and (3) our method, as presented in Section 3.3.1. To fairly compare the second method and ours, we also set $M_0 = 200$ and $M_\Delta = 20$ in our method, as discussed in previous work [1].

Table 3.1: Quantitative comparison of fiducial registration error f_e, target tracking error t_e, and distance between transformed points and assigned centerlines on 15 experiments and their p-values during static phantom validation (Eqs. 3.23, 3.24, and 3.13) (unit: mm).

Methods	Klein et al. [2]			Deguchi et al. [1]			Our method		
Experiments	Error f_e	Distance	Error t_e	Error f_e	Distance	Error t_e	Error f_e	Distance	Error t_e
1	6.66 mm	7.02 mm	6.97 mm	5.66 mm	5.02 mm	4.80 mm	2.06 mm	2.21 mm	1.53 mm
2	8.29 mm	6.97 mm	7.00 mm	9.88 mm	5.20 mm	6.89 mm	5.85 mm	2.60 mm	1.93 mm
3	8.53 mm	6.87 mm	6.67 mm	7.03 mm	5.09 mm	6.23 mm	4.29 mm	2.24 mm	2.23 mm
4	23.1 mm	7.10 mm	7.21 mm	8.56 mm	5.01 mm	4.75 mm	5.02 mm	2.30 mm	1.64 mm
5	11.9 mm	7.36 mm	7.78 mm	6.95 mm	4.78 mm	4.51 mm	3.84 mm	2.16 mm	2.54 mm
6	10.5 mm	7.04 mm	7.96 mm	7.78 mm	5.20 mm	5.03 mm	4.43 mm	2.18 mm	1.47 mm
7	13.8 mm	6.66 mm	8.59 mm	9.25 mm	4.29 mm	5.28 mm	4.57 mm	1.93 mm	3.99 mm
8	6.36 mm	7.69 mm	6.68 mm	9.56 mm	5.68 mm	3.80 mm	8.57 mm	2.67 mm	4.06 mm
9	6.35 mm	7.19 mm	6.07 mm	7.69 mm	5.54 mm	5.78 mm	6.44 mm	2.73 mm	5.34 mm
10	9.72 mm	8.15 mm	7.52 mm	5.64 mm	5.89 mm	5.11 mm	2.82 mm	2.79 mm	2.78 mm
11	11.5 mm	8.10 mm	8.85 mm	7.73 mm	5.05 mm	6.12 mm	4.30 mm	2.46 mm	2.33 mm
12	12.4 mm	8.56 mm	7.81 mm	4.87 mm	5.62 mm	4.37 mm	2.47 mm	2.76 mm	2.42 mm
13	8.80 mm	7.77 mm	8.38 mm	6.93 mm	5.63 mm	5.14 mm	2.91 mm	2.69 mm	2.27 mm
14	9.47 mm	8.23 mm	8.13 mm	7.93 mm	6.11 mm	6.56 mm	3.79 mm	3.14 mm	2.11 mm
15	10.6 mm	7.72 mm	7.07 mm	4.88 mm	5.96 mm	5.46 mm	2.66 mm	3.03 mm	2.17 mm
Average	**10.5 mm**	**7.50 mm**	**7.50 mm**	**7.36 mm**	**5.34 mm**	**5.32 mm**	**4.27 mm**	**2.53 mm**	**2.59 mm**
***P*-value**	**0.43**	**0.95**	**1.00**	**1.14**	**1.14**	**0.82**	**0.78**	**0.87**	**0.40**

To investigate tracking performance of these different approaches, we compute fiducial registration error f_e using markers that were placed on our phantoms:

$$f_e = \frac{1}{N}\sum_{n=1}^{N} ||\mathbf{q}_n^{ct} - {}^{ct}\tilde{\mathcal{T}}_{et}\mathbf{p}_n^{et}||, \tag{3.23}$$

where $\mathbf{q}_n^{ct}$ denotes one marker's position that is manually determined from the CT volume, $\mathbf{p}_n^{et}$ is a measured marker's position using an ET sensor, and ${}^{ct}\tilde{\mathcal{T}}_{et}$ is the final estimate after W iterations in one experiment.

We also evaluate target tracking error of each ET sensor measurement. We performed marker-based registration to obtain ET-CT spatial transformation ${}^{ct}\mathcal{T}_{et}^{*}$, which was assumed to be highly accurate since its fiducial registration error was about 0.8 mm in our experiments. We calculate target tracking error (relative error) t_e by:

$$t_e = \frac{1}{M}\sum_{i=1}^{M} ||{}^{ct}\mathcal{T}_{et}^{*}\mathbf{p}_i - {}^{ct}\tilde{\mathcal{T}}_{et}\mathbf{p}_i||, \tag{3.24}$$

where $\mathbf{p}_i$ is one measured position and ${}^{ct}\tilde{\mathcal{T}}_{et}$ is the final estimate after W iterations in one experiment.

Additionally, we compute the distance between transformed and projected points, ${}^{ct}\tilde{\mathcal{T}}_{et}\tilde{\mathbf{p}}_i$ and $\mathbf{q}_i$, of selected point $\tilde{\mathbf{p}}_i$ of each ET measurement $\mathbf{p}_i$ using ET-CT transformation ${}^{ct}\tilde{\mathcal{T}}_{et}$ (Eq. 3.16). This indicates whether selected point $\tilde{\mathbf{p}}_i$ is closer to centerlines than $\mathbf{p}_i$ after transforming it to the CT volume.

3.3.3 Result

We performed 15 experiments using static and dynamic phantoms respectively to validate our proposed method.

Static Phantom Study

Table 3.1 quantitatively summarizes fiducial registration errors, distances between optimal selected points and assigned centerlines, target tracking errors using different methods. The average fiducial registration and target tracking errors were significantly reduced from at least 7.36 mm and 5.32 mm to 4.27 mm and 2.59 mm, respectively. The average distance also was diminished from 5.34 mm to 2.53 mm.

Figs. 3.16 $\sim$ 3.18 plot the fiducial registration error, distance between transformed points and centerlines, and target tracking error by using different methods that were evaluated

Table 3.2: Quantitative comparison of fiducial registration error f_e, target tracking error t_e, and distance between transformed points and assigned centerlines on 15 experiments and their p-values during dynamic phantom validation (Eqs. 3.23, 3.24, and 3.13) (unit: mm).

Methods	Klein et al. [2]			Deguchi et al. [1]			Our method		
Experiments	Error f_e	Distance	Error t_e	Error f_e	Distance	Error t_e	Error f_e	Distance	Error t_e
1	10.5 mm	9.45 mm	8.59 mm	5.66 mm	6.41 mm	7.49 mm	3.20 mm	4.67 mm	2.65 mm
2	11.3 mm	10.6 mm	9.19 mm	11.1 mm	5.57 mm	10.2 mm	8.23 mm	2.64 mm	4.92 mm
3	8.56 mm	10.5 mm	6.86 mm	9.83 mm	5.68 mm	10.1 mm	8.14 mm	4.53 mm	3.18 mm
4	25.9 mm	9.47 mm	9.42 mm	11.1 mm	5.80 mm	4.97 mm	5.76 mm	5.46 mm	5.47 mm
5	14.7 mm	8.69 mm	8.88 mm	9.12 mm	7.47 mm	6.32 mm	4.62 mm	3.10 mm	5.02 mm
6	13.1 mm	10.4 mm	8.92 mm	9.54 mm	6.92 mm	7.36 mm	5.79 mm	3.98 mm	3.87 mm
7	16.0 mm	8.43 mm	9.57 mm	10.4 mm	7.07 mm	8.03 mm	8.31 mm	4.21 mm	4.68 mm
8	7.23 mm	11.3 mm	7.29 mm	11.6 mm	6.71 mm	6.68 mm	10.1 mm	2.91 mm	4.42 mm
9	9.44 mm	7.33 mm	9.89 mm	10.7 mm	5.57 mm	8.38 mm	7.53 mm	4.71 mm	6.36 mm
10	10.6 mm	10.3 mm	11.3 mm	8.68 mm	8.02 mm	8.01 mm	3.42 mm	5.36 mm	6.22 mm
11	13.0 mm	11.0 mm	12.1 mm	10.0 mm	6.17 mm	7.61 mm	5.89 mm	3.35 mm	5.98 mm
12	16.0 mm	9.27 mm	10.7 mm	7.86 mm	9.41 mm	6.70 mm	3.97 mm	6.11 mm	5.22 mm
13	12.2 mm	9.11 mm	9.08 mm	9.51 mm	9.26 mm	5.61 mm	3.43 mm	6.58 mm	5.17 mm
14	11.1 mm	8.98 mm	9.58 mm	8.43 mm	7.68 mm	6.79 mm	5.53 mm	6.53 mm	3.03 mm
15	11.9 mm	9.01 mm	7.82 mm	6.90 mm	6.06 mm	9.38 mm	3.03 mm	5.06 mm	4.47 mm
Average	**12.8 mm**	**9.59 mm**	**9.28 mm**	**9.36 mm**	**6.92 mm**	**7.58 mm**	**5.80 mm**	**4.61 mm**	**4.71 mm**
***P*-value**	**1.31**	**1.14**	**0.87**	**1.89**	**0.65**	**0.82**	**0.78**	**1.05**	**1.09**

on 15 experiments during static phantom validation. Fig. 3.19 visualizes the distance from transformed points of using different methods to assigned centerlines. Obviously, selecting optimal points and transforming them to CT on the basis of our multiple point strategy is closer to assigned centerlines than other approaches. Fig. 3.20 displays endoscope positions that were estimated by different approaches. Our method still outperforms others.

Dynamic Phantom Study

During dynamic phantom validation, the average fiducial registration and target tracking errors were increased from 4.27 mm and 2.59 mm (errors in static phantom validation) to 5.80 mm and 4.71 mm due to simulated respiratory motion. Tables 3.1 and 3.2 compare the experimental results of using methods of Klein et al. [2], Deguchi et al. [1], and ours. The average distance also was increased from 2.53 mm (distance in static phantom validation) to 4.61 mm. However, our proposed method still performs much better than other two approaches (Table 3.2). Based on the non-parametric statistical hypothesis Wilcoxon signed-rank test, we also computed p-values for the fiducial registration error, the distance, and the target tracking error of using the different methods (Tables 3.1 and 3.2). Because these p-values were large, the null hypothesis cannot be rejected.

Figs. 3.21 $\sim$ 3.23 compare the fiducial registration error, distance between transformed points and centerlines, and target tracking error from different approaches that were evaluated on 15 experiments during dynamic phantom validation. Fig. 3.24 plots endoscope positions of using method of Deguchi et al. [1] and our proposed method against marker-based position estimations. The position estimates of our method are closer to marker-based positions than Deguchi et al. [1]. This implies that our estimates are more accurate than others. Figs. 3.25 and 3.26 illustrate tracking results that were interpreted to virtual rendering images. We use CT images and estimated endoscope positions (tracking results) and perform volume rendering to generate virtual images. We inspect whether virtual rendering images resembles video images. The virtual rendering images of using our method were more similar than others (Figs. 3.25 and 3.26), which implies ours still outperforms other approaches.

3.3.4 Discussion

In general, our proposed method to establish the ET-CT spatial alignment outperforms previous methods. Both fiducial registration and target tracking accuracies were significantly improved during static and dynamic phantom validations. We attribute this improvement to our multiple point strategy.

We designed a model and printed it in 3D to calibrate the spatial relationship between the endoscope center and the ET sensor. Based on this calibrated spatial relationship and each ET sensor measurement (position and orientation), we generate multiple positions or points and select optimal points closest to assigned centerlines. Our strategy fully uses ET sensor measured orientation information in three direction and adaptively choose the closest points. The adaptivity of our method enables optimal points to be used in minimization, significantly diminishing the assumption of manipulating endoscope along the centerlines of body cavities. Also note that the aim of this work is to reduce the assumption of moving the endoscope along the organ's centerline to improve the ET-CT registration and synchronization. We believe that this assumption was successfully diminished, i.e., the distance between the sensor position after transformed to CT and the bronchial centerlines was much reduced, as shown in our experimental results, e.g., Tables 3.1 and 3.2, Fig. 3.17, and Fig. 3.22.

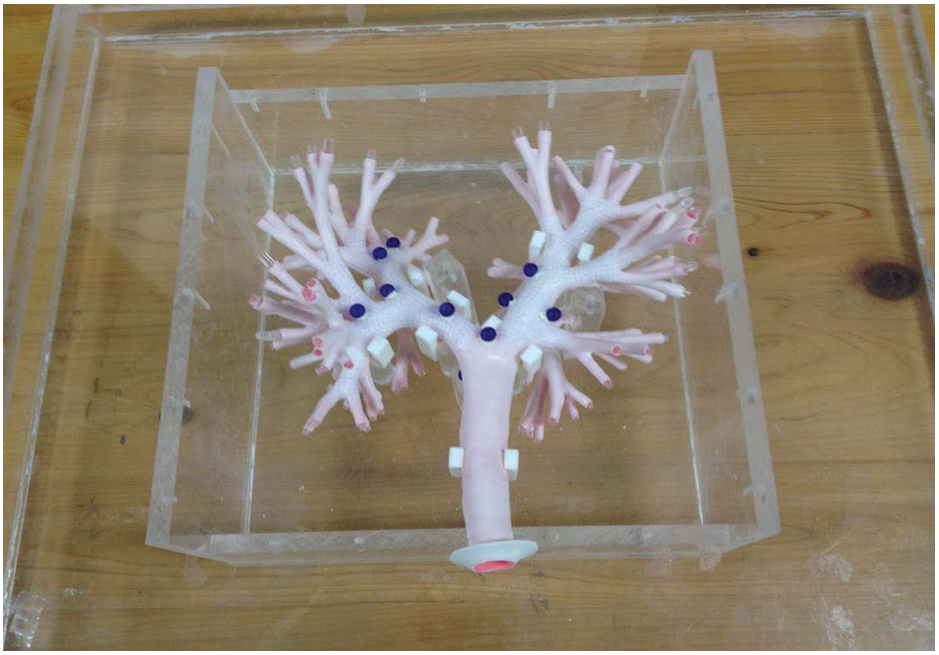
However, our method to align the ET system and the CT volume for endoscope electromagnetic navigation provides limited tracking accuracy, i.e., we still need to further improve the target tracking error of using our method, particularly during dynamic phantom validation. Several potential limitations are clarified as follows. First, patient movements, particularly respiratory motion, usually deteriorates the tracking accuracy, since they might cause inaccurate sensor measurements that may not exactly correspond to the current endoscope position. One future work is to compensate respiratory motion for accurate endoscope tracking. Next, ET systems unavoidably involve static and jitter errors. The static or inherent error of ET systems was about $0.88 \sim 1.4$ mm that were reported in the ET system's specification. The jitter or dynamic error, which remains influenced by the magnetic field distortion of ET systems, is difficult to be corrected. In our future work, we address these errors by a state space model to further improve the tracking accuracy during endoscope electromagnetic navigation. Also note that our proposed method is a rigid registration procedure, i.e., no matter how accurate the ET-CT spatial transformation is determined, the tracking accuracy of using our method limits to respiratory motion and magnetic field distortion, particularly endoscope electromagnetic navigation itself is a noisy, dynamic, and uncertain procedure. Additionally, the endoscope center calibration possibly introduces the registration error that was mainly caused by the ET system's inaccuracy, i.e., the inaccurate measurement of the ET sensor due to static and jitter errors.

Finally, we clarify that our method provides real-time endoscope electromagnetic navigation. The computation of the spatial ET-CT transformation usually takes about 36 – 54 seconds when using two or three thousands of ET sensor measurements in our proposed approach. Once the ET-CT spatial alignment is determined, we temporally track endoscopes in real time since endoscope position and orientation parameters can be computed

by multiplying each ET sensor measurement by the estimated ET-CT spatial transformation matrix.

3.3.5 Conclusions

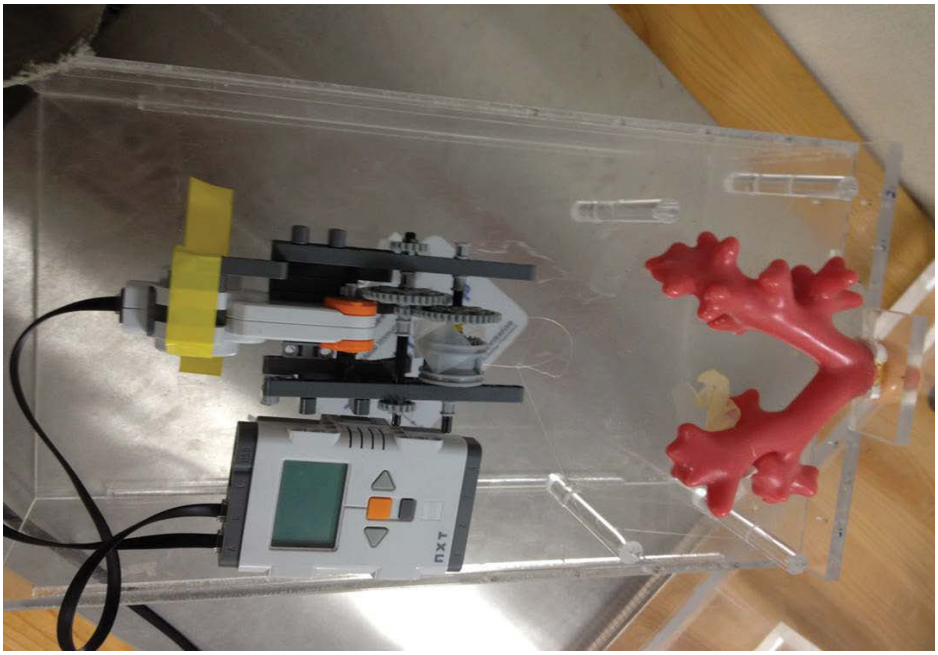
This work proposed an adaptive fiducial-free registration approach to establish the ET-CT spatial alignment for real-time electromagnetically navigated endoscopy. We introduced a multiple point strategy to select optimal points based on ET sensor measured position and orientation information and a procedure of endoscope center calibration to tackle an assumption that an endoscope is operated along the hollow organ's centerlines during endoscopic interventions. Compared to previous methods, the current tracking accuracy of using our proposed adaptive registration approach was significantly improved from at least 5.32 to 2.59 mm in static phantoms validation, as well as from at least 7.58 mm to 4.71 mm in dynamic phantom validation. Future work also includes validation on clinical datasets and application of our proposed method to other flexible registration cases.



(a)

(b)

Figure 3.14: Static phantom (a) and endobronchial ultrasound endoscope with an ET (AURORA) sensor fixed at its distal tip (b).



(a)

(b)

Figure 3.15: Dynamic phantom (a) and endoscope with an ET (Ascension) sensor fixed at its distal tip (b)

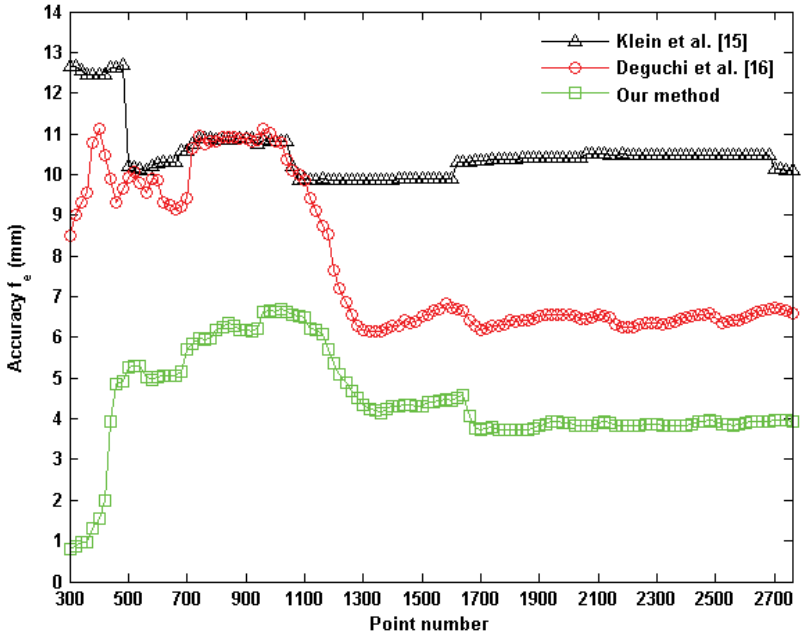


Figure 3.16: Plotted fiducial registration error f_e in static phantom validation (Experiment 6)

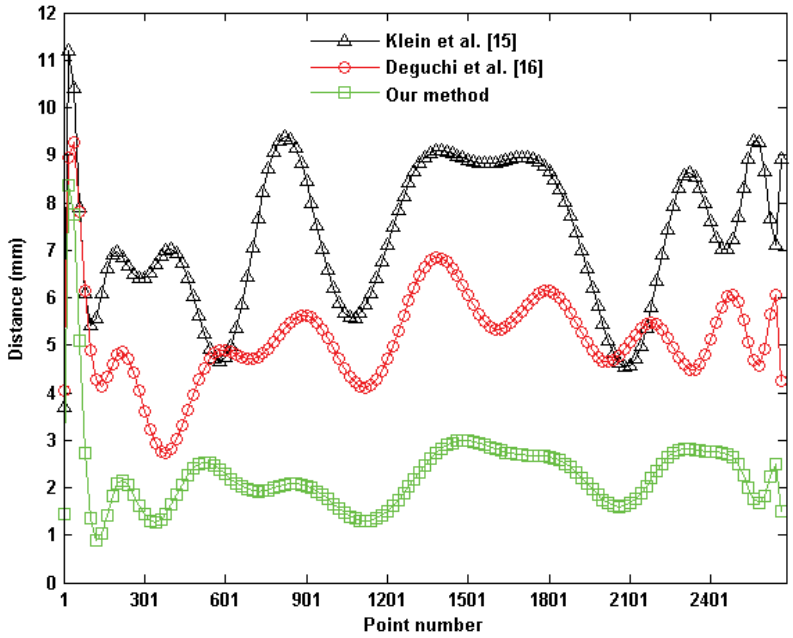


Figure 3.17: Plotted distance to centerlines in static phantom validation (Experiment 4)

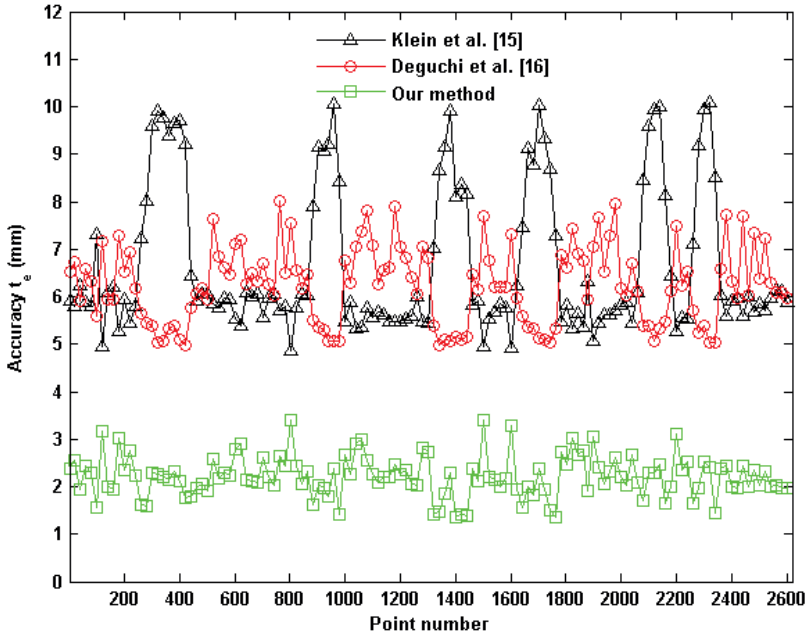


Figure 3.18: Plotted target tracking error t_e in static phantom validation (Experiment 3)

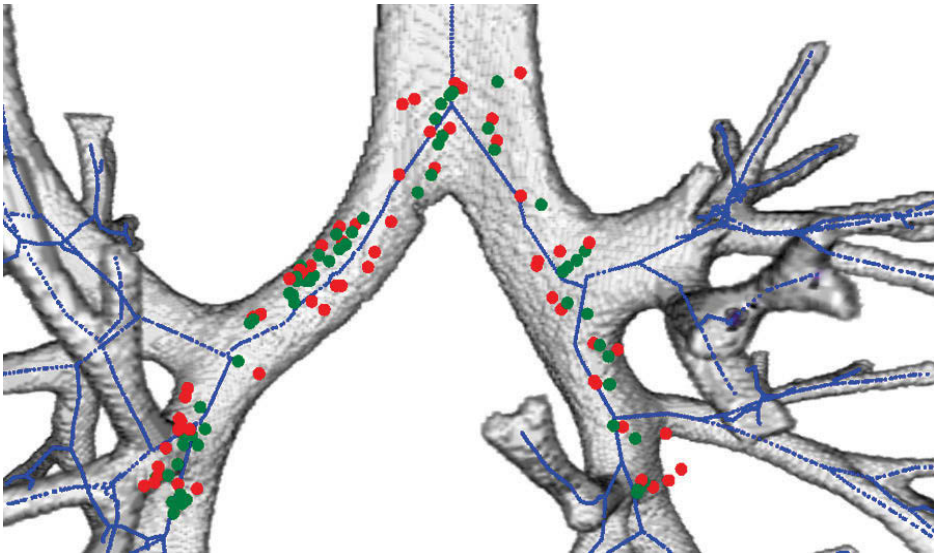


Figure 3.19: Transformed points were plotted along assigned centerlines in static phantom validation (Experiment 14). Points (*green*) transformed by our method were closer to assigned centerlines (*blue*) than points (*red*) transformed by Deguchi et al. [1].

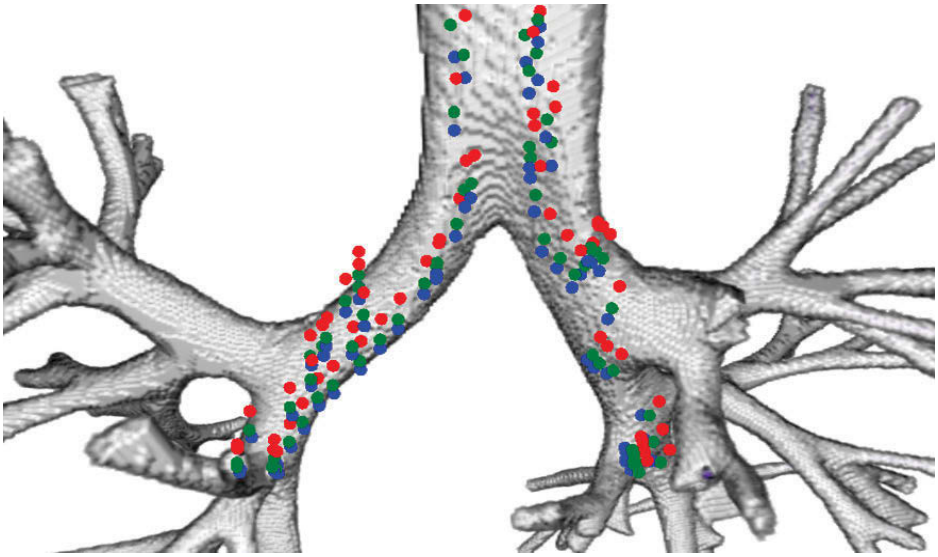


Figure 3.20: Comparison of positions estimated by Deguchi et al. [1] and our method in static phantom validation (Experiment 2). Positions at *blue* points were obtained using marker-based ET-CT transformation ${}^{ct}\mathcal{T}_{et}^{*}$, whose fiducial error was about 0.81 mm. Positions at *green* points estimated by our method were almost overlapping *blue* points which were far from positions at *red* points tracked by Deguchi et al. [1]. The tracking accuracy of our method is significantly better than Deguchi et al. [1].

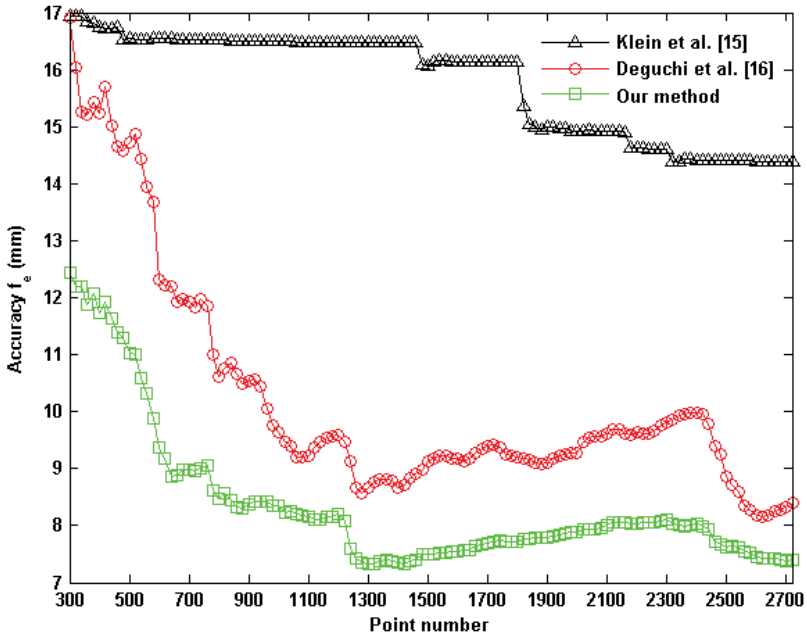


Figure 3.21: Plotted fiducial registration error f_e in dynamic phantom validation (Experiment 7)

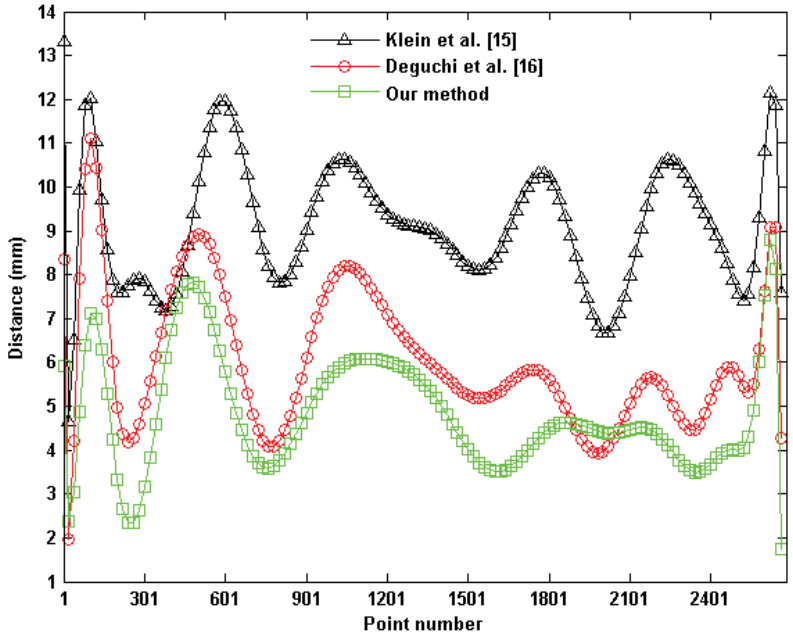


Figure 3.22: Plotted distance to centerlines in dynamic phantom validation (Experiment 15)

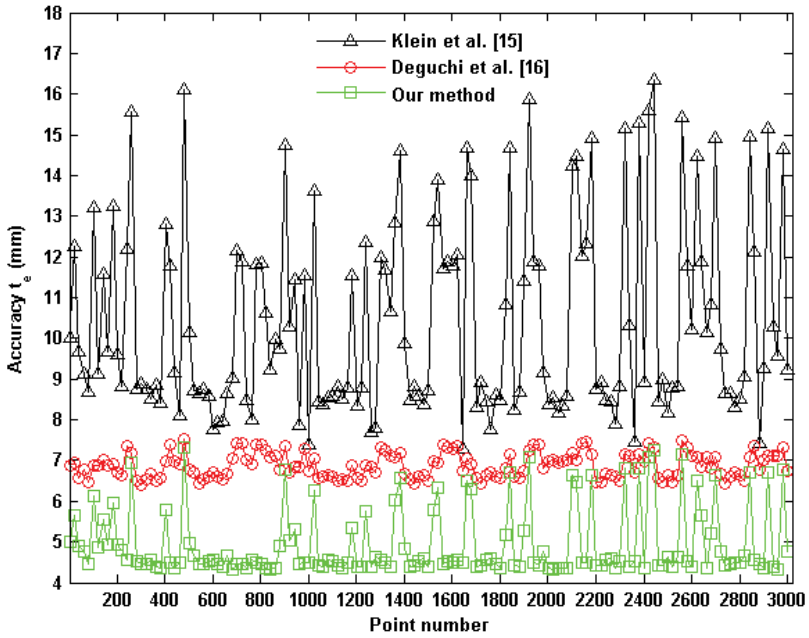
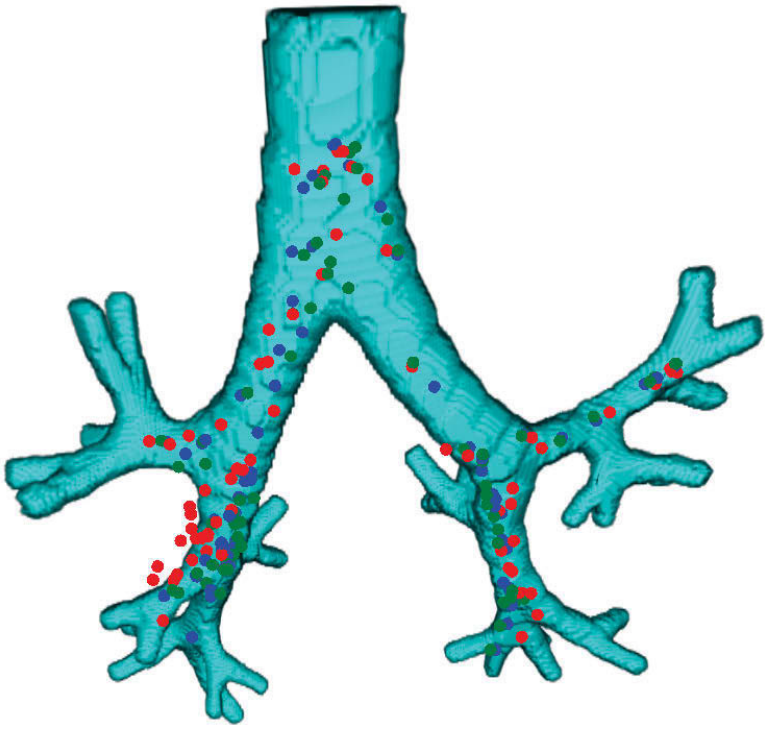
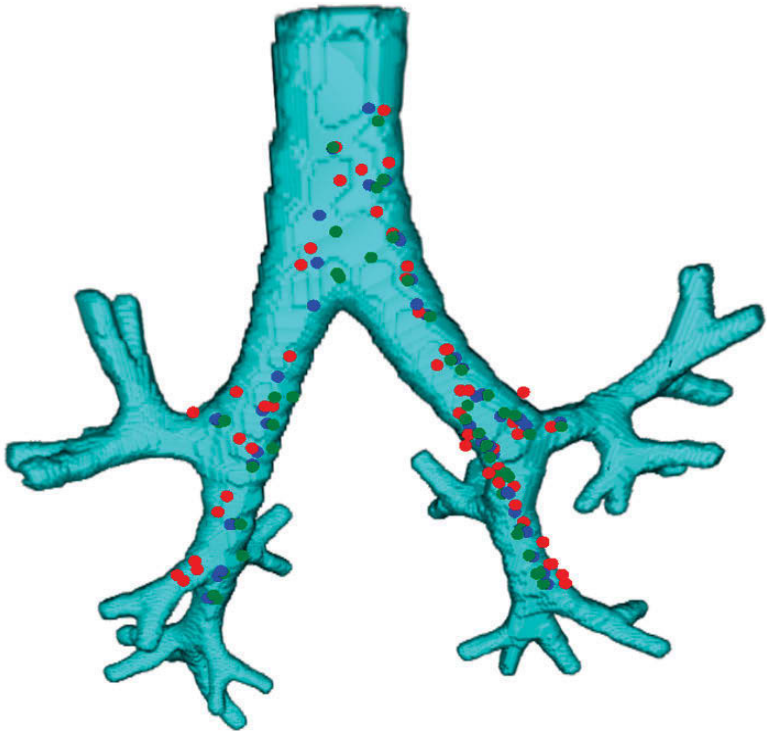


Figure 3.23: Plotted target tracking error t_e (Experiment 12) in dynamic phantom validation



(a) Comparison of positions (Experiment 5)



(b) Comparison of positions (Experiment 8)

Figure 3.24: Examples of comparison of tracking results of using different methods on Experiments 5 and 8 during dynamic phantom validation: Plotted endoscope positions estimated by Deguchi et al. [1] (*red* points) and our method (*green* points) against marker-based estimations (*blue* points).

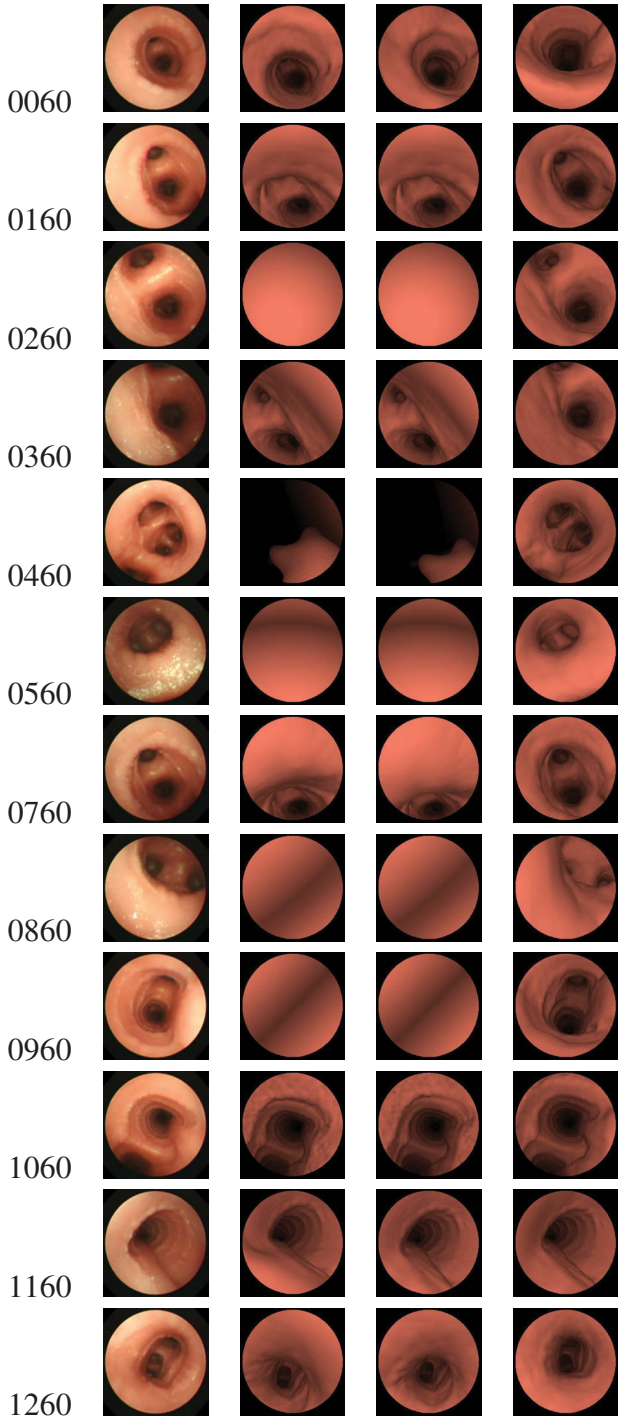


Figure 3.25: Examples of comparison of tracking results of using different methods on Experiments 5 during dynamic phantom validation. First column show frame numbers selected uniformly every 100 frames, and second column displays corresponding videos images, and other columns show virtual images generated from tracking results using different methods of Klein et al. [2], Deguchi et al. [1], and ours. The proposed method performs much better than other two approaches.

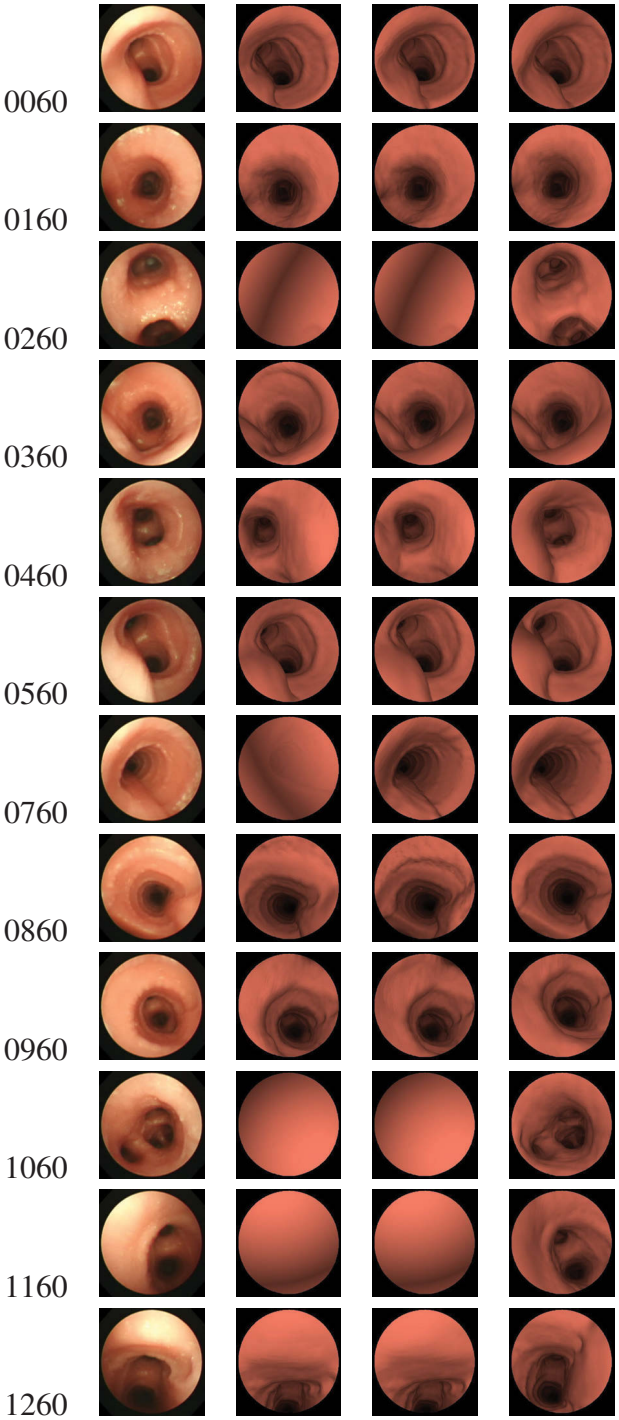


Figure 3.26: Examples of comparison of tracking results of using different methods on Experiments 5 during dynamic phantom validation. TThe first column show frame numbers selected uniformly every 100 frames, and the second column displays corresponding videos images, and other rows show virtual images generated from tracking results using different methods of Klein et al. [2], Deguchi et al. [1], and ours. The proposed method performs much better than other two approaches.

Chapter 4

Enhanced particle swarm optimization

This chapter presents a new framework on the basis of an enhanced particle swarm optimization method to effectively fuse these information for accurate and continuous endoscope localization, which corresponds to our journal paper titled "Robust electromagnetically guided endoscopic procedure using enhanced particle swarm optimization for multimodal information fusion" that was published in Medical Physics.

We use the particle swarm optimization method, which is one of stochastic evolutionary computation algorithms, to effectively fuse the multimodal information including preoperative information (i.e., computed tomography images) as a frame of reference, endoscopic camera videos, and positional sensor measurements (i.e., electromagnetic sensor outputs). Since the evolutionary computation method usually limits its possible premature convergence and evolutionary factors, we introduce the current (endoscopic camera and electromagnetic sensor's) observation to boost the particle swarm optimization and also adaptively update evolutionary parameters in accordance with spatial constraints and the current observation, resulting in advantageous performance in the enhanced algorithm. Our proposed framework greatly reduced the guidance errors from (4.3, 7.8) to (3.0 mm, 5.6°), compared to state-of-the-art methods.

4.1 Methods and Materials

Electromagnetically guided endoscopy seeks to accurately and continuously determine the endoscope location or its 6DoF motion parameters including position and orientation in a reference frame (e.g., the CT image coordinate system). The core of our endoscopic guidance framework is to use the EPSO algorithm to integrate multimodal information to precisely navigate the endoscope. Before discussing EPSO, we slightly introduce different sensory information involved in the endoscopic guidance.

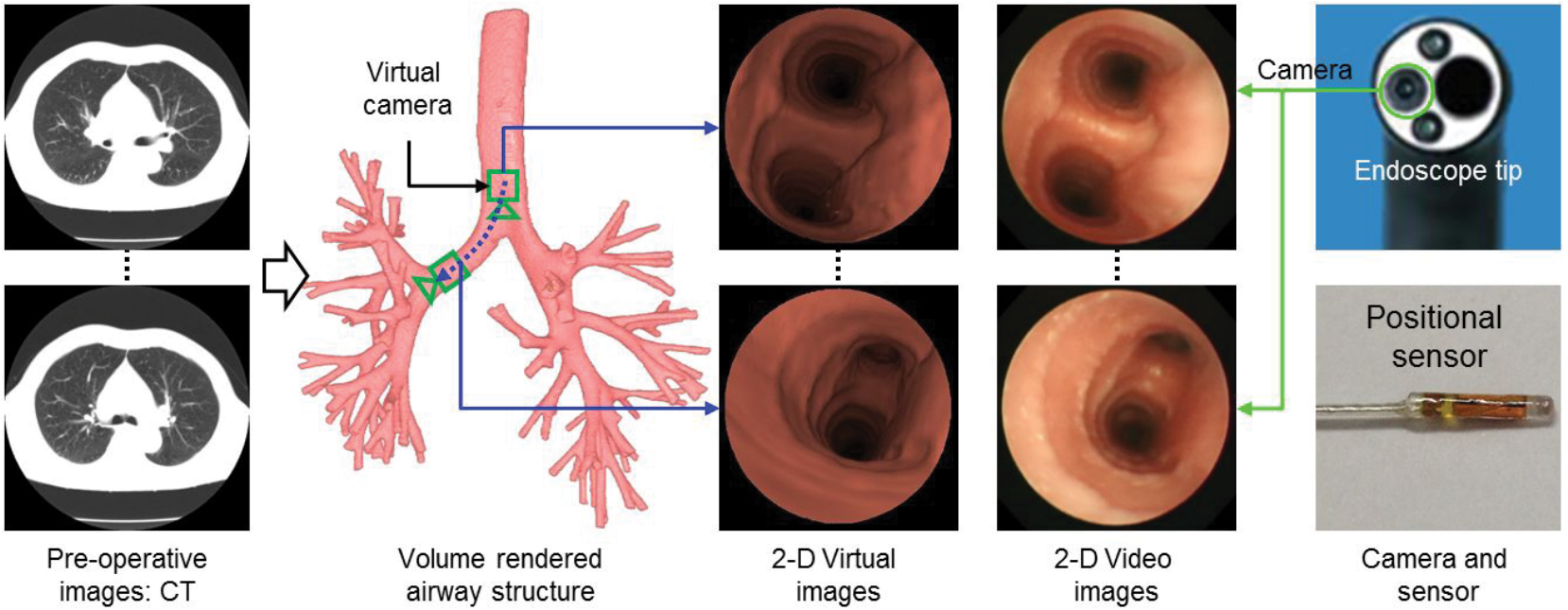


Figure 4.1: Multimodal information including the CT images used to generate 2-D virtual rendering images, endoscopic video images, and positional sensor measurements involved in an electromagnetically guided endoscopic procedure.

4.1.1 Multimodal information

Pre-operative information

Before performing endoscopic examinations, physicians usually use 3-D imaging devices such as CT scanners to acquire pre-operative images. We use the CT images to generate 2-D virtual images by volume rendering techniques [38] (Fig. 4.1).

Endoscopic video sequences

The endoscopic camera is directly inserted to observe organ interior structures. Endoscopic video images display sufficient organ interior surface information (Fig. 4.1). However, the camera only provides 2-D live video images without depth information to target regions. Hence, we require to integrate 2-D endoscopic images with 3-D CT images to guide endoscopic procedures.

Positional sensor measurements

Ultra-miniature sensors, i.e, EM sensors (Fig. 4.1), which are externally attached to the endoscope's distal tip surface or its working channel, are increasingly used to measure the endoscopic camera movement information. However, the sensor measurements might be inaccurate due to tissue deformation, e.g., respiration.

4.1.2 Particle Swarm Optimization

The particle swarm optimization (PSO) that was originally proposed by Kennedy and Eberhart [44] has been increasingly applied as an optimization technique to address complex problems [45, 9, 46]. The algorithm simulates natural and biological behaviors such as birds flocking and fish schooling to find optimal solutions in nonlinear and high-dimensional spaces. Moreover, one of most attractive aspects of PSO is able to deal with nonlinear, non-differentiable, and multimodal optimization problems by dynamically interacting all particles in a similar analogy with the *cognitive* and *social* properties of populations [47]. However, the standard PSO method limits to its evolutionary parameters, particularly it does not take the current observation information into consideration when propagating the particle swarm to the new state. These limitations potentially result in a premature convergence. This work first proposes an accurate and robust electromagnetically endoscopic guidance framework that uses an enhanced particle swarm optimization (EPSO) algorithm to address these limitations to improve the PSO's performance.

The PSO algorithm is a population-based stochastic optimization technique. It seeks to propagate the population with a number of particles to approximate the optimal solution. Suppose that we generate a population of particles $\mathscr{P}_j = \{(\mathbf{p}_{i,j}, f(\mathbf{p}_{i,j}), \gamma_{i,j})\}_{i=1}^{M}$ at j-th iteration, where $f(\mathbf{p}_{i,j})$ is the particle fitness value, $\gamma_{i,j}$ is a weight on the basis of spatial distribution constraint, M is the particle number, and $j = 1, 2, \ldots N$, N is the iteration number. Our proposed EPSO algorithm is to update these particles with several steps: (1) randomization, (2) propagation, (3) evolutionary parameters analysis, and (4) population update, as discussed as follows.

4.1.3 Enhanced particle swarm optimization

Population randomization

We first initialize particle $\mathbf{p}_{0,0}$ and $f(\mathbf{p}_{0,0}) = \gamma_{0,0} = 1/M$ before iterations. We perform a randomization only once to increase the diversity of particles and avoid the particle impoverishment. Particle $\mathbf{p}_{0,0}$ is randomized by using the Gaussian propagation model, and we obtain particle state $\mathbf{x}_{0,0}$:

$$\mathbf{x}_{0,0} = \mathscr{G}(\mathbf{p}_{0,0}, \mu\Lambda), \tag{4.1}$$

where $\mathscr{G}$ is the transform function that is used to add changeable part $\mu\Lambda$ to particle $\mathbf{p}_{0,0}$ and obtain new particle $\mathbf{x}_{0,0}$, μ is a Gaussian distribution random number: $\mu \sim \mathscr{N}(0,1)$, and Λ is a pre-determined constant vector. Note that the randomization for particle diversification does not perform a *resampling process*, as particle filter methods do [8], since the local best particles provide compact samples for propagation [48].

Population propagation

After the randomization, we obtain particles $\{\mathbf{x}_{i,j}\}_{i=1}^{M}$ at j-th iteration and propagate $\mathbf{x}_{i,j}$ to $\mathbf{x}_{i,j+1}$ at iteration j+1 on the basis of velocity $\mathbf{v}_{i,j}$ and inertia weight ω that is used for deciding how much $\mathbf{v}_{i,j}$ to be preserved in $\mathbf{v}_{i,j+1}$ (Fig. 4.2):

$$\mathbf{v}_{i,j+1} = \omega\mathbf{v}_{i,j} + \underbrace{\lambda_1\eta_1(\mathbf{p}_{i,j} - \mathbf{x}_{i,j})}_{\Delta_p} + \underbrace{\lambda_2\eta_2(\mathbf{g}_{i,j} - \mathbf{x}_{i,j})}_{\Delta_g}, \tag{4.2}$$

$$\mathbf{x}_{i,j+1} = \mathbf{x}_{i,j} + \mathbf{v}_{i,j+1}, \tag{4.3}$$

where λ_1 and λ_2 are acceleration constants and η_1 and η_2 are randomly generated from the uniform distribution with interval [0.0 1.0]. Particles $\mathbf{p}_{i,j}$ (for the local individual best) and

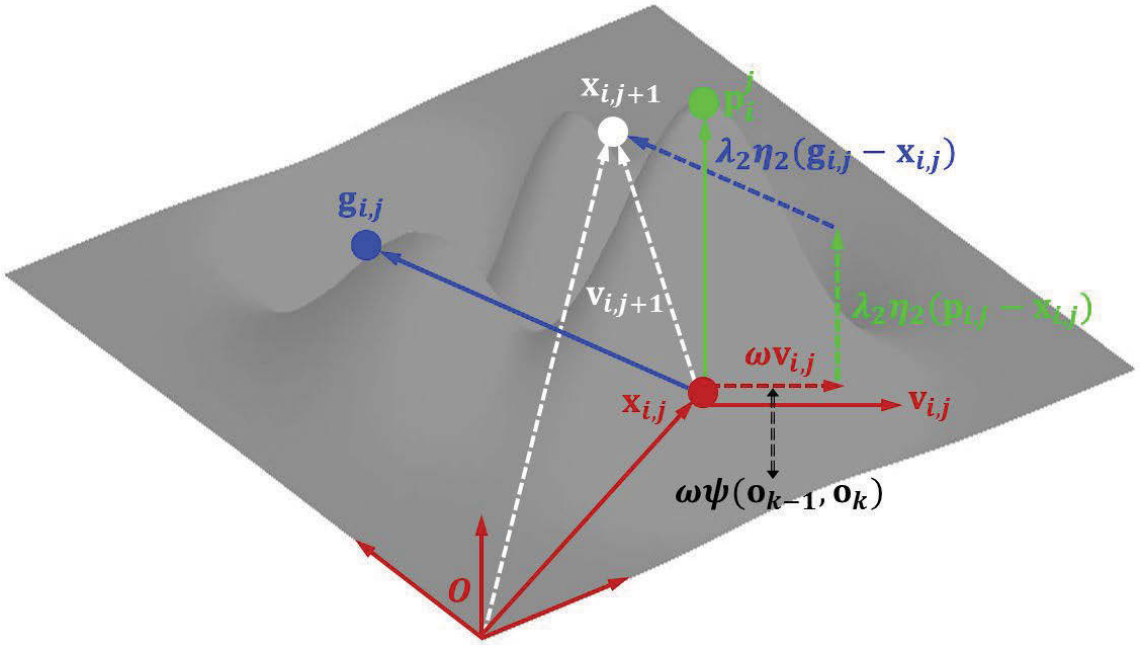


Figure 4.2: Population propagation to obtain new particle states. The *red*, *green*, and *blue* points (or solid lines) denote the current particle $\mathbf{x}_{i,j}$, the local individual best particle $\mathbf{p}_{i,j}$, and the global all best $\mathbf{g}_{i,j}$ at iteration j, respectively. The *red*, *green*, and *blue* dash lines display vectors $\omega\mathbf{v}_{i,j}$, $\lambda_1\eta_1(\mathbf{p}_{i,j} - \mathbf{x}_{i,j})$, and $\lambda_2\eta_2(\mathbf{g}_{i,j} - \mathbf{x}_{i,j})$ (Eq. 4.2) to determine vector $\mathbf{v}_{i,j+1}$ (the short *white* dash line) combined with $\mathbf{x}_{i,j}$ to compute new particle $\mathbf{x}_{i,j+1}$ (the long *white* dash line) in Eq. 4.3. Vector $\omega\mathbf{v}_{i,j}$ depends on the current observation (Eq. 4.4).

$\mathbf{g}_{i,j}$ (for the global all best) are the best state found by particle i so far and the best state found by the whole swarm so far.

The PSO's performance depends on this particle propagation (Eq. 4.3), where velocity $\mathbf{v}_{i,j}$ plays an important role since it aims at keeping all the particles approaching to the optimal solution. It is difficult to accurately determine $\mathbf{v}_{i,j}$. In the standard PSO algorithm, it is empirically determined within a reasonable range, which usually loses the relative continuity between two consecutive times $k-1$ and k from iterations j to $j+1$. PSO might get trapped in a local minima convergence. To address this problem, our idea is to integrate current observation at time k to compute velocity $\mathbf{v}_{i,j}$:

$$\mathbf{v}_{i,j} = \Phi(\mathbf{o}_{k-1}, \mathbf{o}_k), \tag{4.4}$$

where function $\Phi(\cdot)$ is to compute the relative continuity between previous and current observations $\mathbf{o}_{k-1}$ and $\mathbf{o}_k$.

Now velocity $\mathbf{v}_{i,j+1}$ can be deterministically updated by

$$\mathbf{v}_{i,j+1} = \omega \underbrace{\Phi(\mathbf{o}_{k-1}, \mathbf{o}_k)}_{Observation} + \Delta_p + \Delta_g, \tag{4.5}$$

which is effective to control the population's propagation.

Evolutionary parameters analysis

Evolutionary parameters λ_1, λ_2, and ω heavily influence the optimization performance. Most of current improved PSO algorithms do not consider the relative or temporal continuity and spatial constraint information. This might result in a lack of systematic treatment of evolutionary states and expose PSO to a dangerous level of swarm explosion and divergence. To handle this limitation, we here adaptively control acceleration factors λ_1 and λ_2 relative to the particle's fitness value:

$$\lambda_1 = \frac{2f(\mathbf{p}_{i,j})}{f(\mathbf{p}_{i,j}) + f(\mathbf{g}_{i,j})}, \quad \lambda_2 = \frac{2f(\mathbf{g}_{i,j})}{f(\mathbf{p}_{i,j}) + f(\mathbf{g}_{i,j})}, \tag{4.6}$$

where fitness $f(\cdot)$ is defined as observation probability $\pi(\cdot)$ of each particle relative to current observation $\mathbf{o}_k$:

$$f(\mathbf{p}_{i,j}) = \pi(\mathbf{o}_k|\mathbf{p}_{i,j}), \quad f(\mathbf{g}_{i,j}) = \pi(\mathbf{o}_k|\mathbf{g}_{i,j}). \tag{4.7}$$

For adaptively calculating ω, we utilize both fitness value $f(\mathbf{x}_{i,j}) \in [0.0\ 1.0]$ and particle spatial distribution information $\gamma_{i,j}$ among all the particles in the population. We first compute average distance $d_{i,j}$ from one particle to all other particles:

$$d_{i,j} = \frac{1}{M-1} \sum_{i=1, i\neq i'}^{M} \left\| \mathbf{x}_{i,j} - \mathbf{x}_{i',j} \right\|. \tag{4.8}$$

where symbol $||\cdot||$ means the Euclidean norm. After finding the largest distance d_{max} and the smallest distance d_{min} from $\{d_{i,j}\}_{i=1}^{M}$, we normalize distance $d_{i,g}$ between one particle and the global best particle and obtain $\gamma_{i,j}$ for each particle:

$$\gamma_{i,j} = (d_{i,g} - d_{min})/(d_{max} - d_{min}), \quad \gamma_{i,j} \in [0.0\ 1.0]. \tag{4.9}$$

Finally, since inertia weight ω was suggested within the interval [0.4 0.9] for weighting the global and the local searching abilities [47], we can adaptively calculate it by

$$\omega(f(\mathbf{x}_{i,j}),\gamma_{i,j}) = \frac{2}{2+3\exp(-1.28(f(\mathbf{x}_{i,j})+\gamma_{i,j}))}, \tag{4.10}$$

which shows a novel strategy to automatically control ω in our proposed EPSO algorithm. Note that spatial constraint $\gamma_{i,j}$ uses the particle distribution information to limit the particle to be moved far from the optimal solution and enhances the exploitation ability of each particle in the population.

Population update

After the (j+1)-th iteration, we obtain population $\mathscr{P}_{j+1}$ where $\mathbf{p}_{i,j+1}$ and $\mathbf{g}_{i,j+1}$ are updated using fitness $f(\mathbf{x}_{i,j+1})$:

$$\mathbf{p}_{i,j+1} = \begin{cases} \mathbf{x}_{i,j+1} & \text{if} f(\mathbf{x}_{i,j+1}) > f(\mathbf{p}_{i,j}) \\ \mathbf{p}_{i,j} & \text{otherwise} \end{cases}, \tag{4.11}$$

$$\mathbf{g}_{i,j+1} = \arg \max_{\mathbf{p}_{i,j+1} \in \mathscr{P}_{j+1}} f(\mathbf{p}_{i,j+1}). \tag{4.12}$$

Based on Eqs. 4.1∼4.12, all the particles are updated iteratively to approximate the optimal solution during our EPSO procedure.

4.1.4 Application to endoscopic guidance

This section applies the EPSO algorithm to fuse the multimodal information for guiding the endoscopic procedure. Fig. 4.3 shows the flowchart of our EPSO method that integrates the multimodal information for the robust electromagnetically endoscopic guidance during interventions. We first interpolate EM sensor measurements.

Sensor measurement interpolation

Since the endoscope movement is spatially continuous, we perform the Catmull-Rom spline interpolation for endoscope position $\mathbf{t}_k$ at time k (or at frame k) and the spherical linear interpolation (SLERP) for endoscope orientation that is described by quaternion $\mathbf{q}_k$ to smoothly approximate endoscope real movement [49]. Position $\mathbf{t}_k$ is interpolated by:

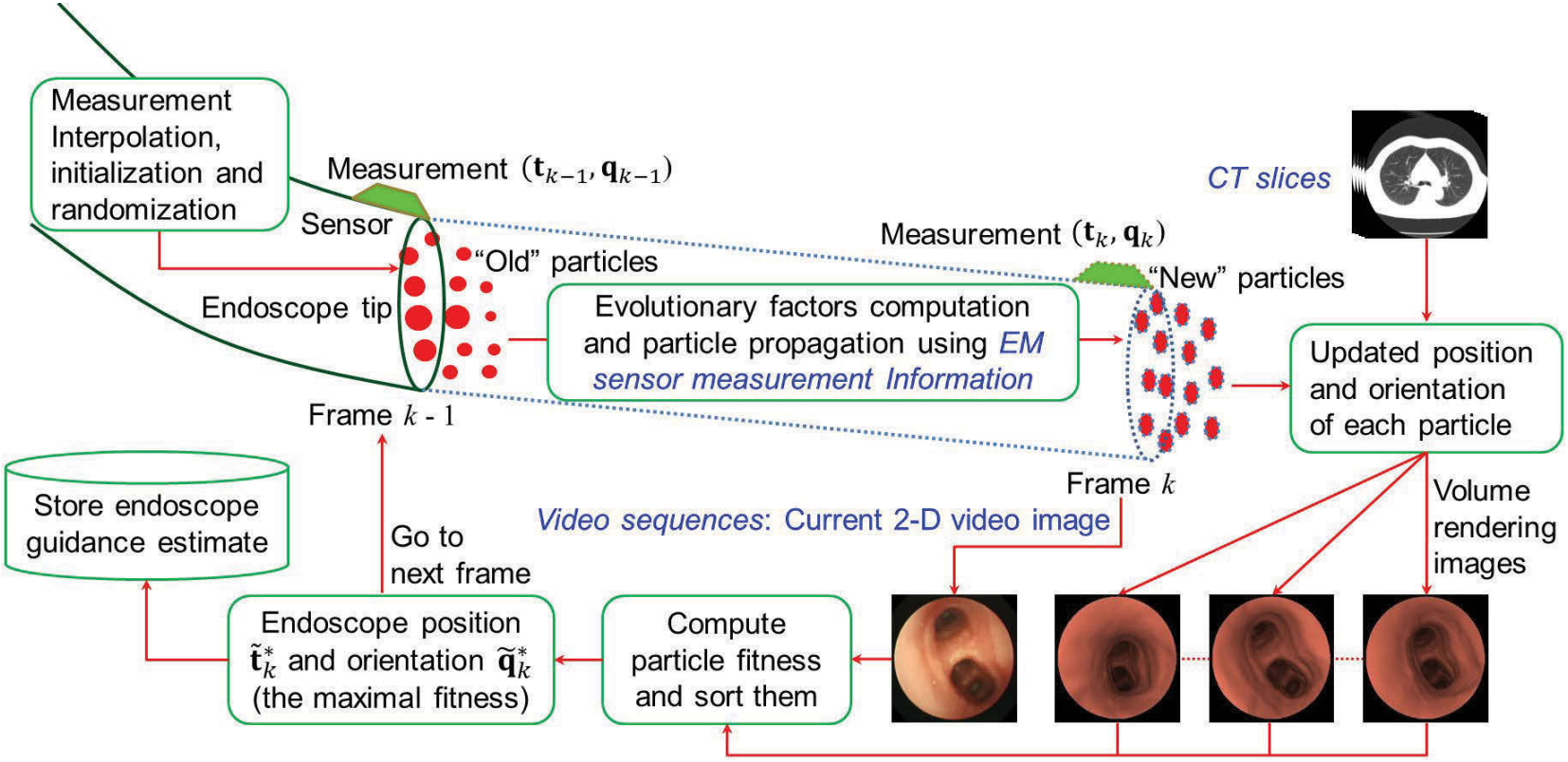


Figure 4.3: Our endoscopic guidance framework using EPSO for the multimodal information integration.

$$\mathbf{t}_k = \Omega \begin{pmatrix} 0 & 1 & 0 & 0 \\ -\beta & 0 & \beta & 0 \\ 2\beta & \beta-3 & 3-2\beta & -\beta \\ -\beta & 2-\beta & \beta-2 & \beta \end{pmatrix} \begin{pmatrix} \mathbf{t}_{v-1} \\ \mathbf{t}_v \\ \mathbf{t}_{v+1} \\ \mathbf{t}_{v+2} \end{pmatrix}, \tag{4.13}$$

where $\Omega = (1\ \alpha\ \alpha^2\ \alpha^3)$, the interpolation ratio $\alpha = k/u - \lfloor k/u \rfloor$, $\mathbf{t}_{v-1} \cdots \mathbf{t}_{v+2}$ are positions of continuous controlled points ($v = \lfloor k/u \rfloor$, floor operator $\lfloor \cdot \rfloor$, time spacing u), and tension parameter β impacts on the curvature at the control points (usually, $\beta = 0.5$). Quaternion $\mathbf{q}_k$ is interpolated by:

$$\mathbf{q}_k = \begin{cases} \frac{\sin(1-\alpha)\phi}{\sin\phi}\mathbf{q}_v + \frac{\sin\alpha\phi}{\sin\phi}\mathbf{q}_{v+1} & \phi \geq 0 \\ \frac{\sin(1-\alpha)\phi}{\sin\phi}\mathbf{q}_v - \frac{\sin\alpha\phi}{\sin\phi}\mathbf{q}_{v+1} & \phi < 0 \end{cases}, \tag{4.14}$$

$$\phi = \arccos \frac{< \mathbf{q}_v,\ \mathbf{q}_{v+1} >}{\| \mathbf{q}_v \| \| \mathbf{q}_{v+1} \|}, \tag{4.15}$$

where $<,>$ is the dot product and $\| \cdot \|$ is the Euclidean norm.

Since a population of particles in the EPSO algorithm denotes the potential solutions in a dynamic system, we define the i-th particle $\mathbf{p}_{i,j}$ at iteration j as a seven-dimensional (7-D) vector with respect to the endoscopic camera 6DoF pose with position $\tilde{\mathbf{t}}_k$ and orientation (quaternion) $\tilde{\mathbf{q}}_k$ at time k:

$$\mathbf{p}_{i,j} = \begin{pmatrix} \tilde{\mathbf{t}}_k & \tilde{\mathbf{q}}_k \end{pmatrix}_{7\times 1}, \tag{4.16}$$

where $\tilde{\mathbf{t}}_k = (\tilde{t}_k^x, \tilde{t}_k^y, \tilde{t}_k^z)_{3\times 1}$ and $\tilde{\mathbf{q}}_k = (\tilde{q}_k^0, \tilde{q}_k^1, \tilde{q}_k^2, \tilde{q}_k^3)_{4\times 1}$.

Particle $\mathbf{p}_{0,0}$ is initialized by the first sensor measurement: $\mathbf{p}_{0,0} = (\mathbf{t}_0\ \mathbf{q}_0)_{7\times 1}$. So, the endoscopic camera motion is parameterized as a 7-D vector with the position and orientation.

EPSO-based guidance framework

The EPSO algorithm performs several steps, as discussed in Section 4.1.3, to fuse endoscopic video images, the CT images, and the EM sensor measurements. It estimates the endoscope's 6DoF position and orientation to locate it during endoscopic examinations.

Let $\mathbf{I}_k$ and $\mathbf{I}_V$ be the k-th endoscopic camera image and 2-D virtual rendering image generated from the CT images, respectively. In our endoscopic guidance, we have two live observations: endoscopic camera image $\mathbf{I}_k$ and EM sensor measurement $\mathbf{m}_k = (\mathbf{t}_k, \mathbf{q}_k)$. Based on current EM sensor observation $\mathbf{m}_k$, the velocity in Eq. 4.5 can be calculated by

$$\mathbf{v}_{i,j+1} = \omega \underbrace{\Phi(\mathbf{m}_{k-1}, \mathbf{m}_k)}_{Observation} + \Delta_p + \Delta_g. \tag{4.17}$$

Based current endoscopic camera observation $\mathbf{I}_k$, we define the fitness of particle $\mathbf{p}_k^{i,j}$ in accordance with Eq. 4.7:

$$f(\mathbf{p}_{i,j}) = \pi(\mathbf{I}_k|\mathbf{I}_V(\mathbf{p}_{i,j})), \tag{4.18}$$

where $\mathbf{I}_V(\mathbf{p}_{i,j})$ denotes the virtual image corresponding to particle $\mathbf{p}_{i,j}$. It is natural to define probability $\pi(\mathbf{I}_k|\mathbf{I}_V(\mathbf{p}_{i,j}))$ as the similarity between images $\mathbf{I}_k$ and $\mathbf{I}_V(\mathbf{p}_{i,j})$:

$$\pi(\mathbf{I}_k|\mathbf{I}_V(\mathbf{p}_{i,j})) = S(\mathbf{I}_k, \mathbf{I}_V(\mathbf{p}_{i,j})). \tag{4.19}$$

Before computing similarity $S(\mathbf{I}_k, \mathbf{I}_V(\mathbf{p}_{i,j}))$, we detect the specific patches with structural information from the camera video image $\mathbf{I}_k$ (Fig 4.4). The extraction of the structural patches is to calculate the pixel color information of saturation and lightness in the hue-saturation-lightness (HSL) color model [24]. We pre-define two thresholds for the pixel saturation and lightness and discard the patches without structural information. Based on the mean square error measure [24, 8], we calculate similarity $S(\mathbf{I}_k, \mathbf{I}_V(\mathbf{p}_{i,j}))$ in the detected patches from endoscopic camera image $\mathbf{I}_k$:

$$S(\mathbf{I}_k, \mathbf{I}_V(\mathbf{p}_{i,j})) = \frac{1}{U}\sum_{U}\frac{1}{W}\sum_{(a,b)\in W} M, \tag{4.20}$$

$$M = \left((\mathbf{I}_k(a,b) - \bar{\mathbf{I}}_k) - (\mathbf{I}_V(a,b) - \bar{\mathbf{I}}_V)\right)^2 \tag{4.21}$$

where U is the number of the detected patches, W is the number of pixels in one patch, $\bar{\mathbf{I}}_k$ and $\bar{\mathbf{I}}_V$ are the average intensities of all the detected patches from images $\mathbf{I}_k$ and $\mathbf{I}_V$.

We perform N iterations in EPSO at frame k and obtain best solution set $\mathscr{B}_k = \{\mathbf{g}_{i,j+1}\}_{j=1}^{N}$ by storing global best solution $\mathbf{g}_{i,j+1}$ at each iteration. We select optimal solution $\mathbf{g}_k^*$ from $\mathscr{B}_k$ on the basis of fitness $f(\mathbf{g}_{i,j+1})$ (Eqs. 4.18$\sim$4.21). Eventually, the output of EPSO, which fuses the CT images, endoscopic videos, and EM sensor measurements for the endoscopic guidance, is the endoscope's position $\tilde{\mathbf{t}}_k^*$ and orientation $\tilde{\mathbf{q}}_k^*$:

$$(\tilde{\mathbf{t}}_k^* \quad \tilde{\mathbf{q}}_k^*) \longleftrightarrow \mathbf{g}_k^* = \arg\max_{\mathbf{g}_{i,j+1}\in\mathscr{B}_k} f(\mathbf{g}_{i,j+1}). \tag{4.22}$$

4.1.5 Validation

Information acquisition

We validated our proposed method on a dynamic phantom with an adjustable motion: $0 \sim 24$ mm. The CT images of our phantom were acquired with parameters of $512\times512\times1021$ voxels and $0.68\times0.68\times0.5$ mm^3. A 3-D Guidance medSAFE tracker (Ascension Technology

Corporation, USA) was used as an EMT system, which includes a 9-coil at flat-type transmitter as a magnetic field generator and an EMT sensor (1.5 mm, 6DoF). Endoscopic video images of size 362×370 pixels were collected at 30 frames per second using an endoscope (BF Type P260F, Olympus, Tokyo).

We investigate five approaches for endoscope location: (1) M1: reported by Schwarz et al. [4], (2) M2: a hybrid method by Mori et al. [5], (3) M3, a method to selectively use EMT sensor measurements by Luo et al. [6], (4) M4, a sequential Monte Carlo-based method introduced by Luo et al. [8], (5) M5, our proposed method, as discussed in Section 4.1.3. All the methods were evaluated on a laptop with Windows 7 Professional 64-Bit System installed with Microsoft C++, 16.0-GB Memory, and Processor Intel(R) Core(TM) i7 CPU×8.

Evaluation criteria

To evaluate the guidance accuracy of different methods, we generated five ground truth datasets (GTDs) by manually adjusting the position and orientation of the virtual camera to qualitatively register the video and virtual camera viewing points by hand. Two observers independently and repeatedly collected these GTDs. We clarify that intra-observer consistency was 1.81 mm and 5.9°, 1.76 mm and 4.9°, and 1.93 mm and 4.8° from the two observers, respectively; inter-observer consistency was 1.71 mm and 5.6°. Based on GTDs, position and orientation errors, ζ and ψ, are computed by:

$$\zeta = ||\mathbf{t} - \mathbf{t}_G||, \quad \varepsilon = \psi(\mathbf{q}, \mathbf{q}_G), \tag{4.23}$$

where $\mathbf{t}$ and $\mathbf{t}_G$ are the estimated and ground-truth positions, and $\mathbf{q}$ and $\mathbf{q}_G$ are the estimated and ground-truth orientations.

Even though the accuracy and runtime of the endoscopic guidance are much significant for clinical applications, it is necessary to evaluate the smoothness of the guidance procedure, since the endoscope movement is usually a continuous procedure where the camera trajectory should be a smooth curve. Smooth guidance implies little jitter or jump errors might be involved in the endoscope movement estimation. We define the guidance smoothness as the average inter-frame distance for a series of endoscope position $\mathbf{t}_k$ and orientation $\mathbf{q}_k$ that are estimated by the five approaches:

$$\frac{1}{K-1}\sum_{k=1}^{K-1} ||\mathbf{t}_{k+1} - \mathbf{t}_k||, \quad \frac{1}{K-1}\sum_{k=1}^{K-1} \psi(\mathbf{q}_{k+1}, \mathbf{q}_k). \tag{4.24}$$

To evaluate the visual quality of 2-D virtual rendering images that were generated from the estimated position and orientation of the five methods, we compute normalized cross

correlation (NCC) coefficient $C_{k,V}$ between each video image $\mathbf{I}_k$ and its estimated virtual image $\mathbf{I}_V$ at time k in an experiment. The average visual quality $\mathscr{Q}$ of one endoscopic video sequence for testing different methods can be defined as:

$$\mathscr{Q} = \frac{1}{K}\sum_{k=1}^{K}\frac{(C_{k,V}+1)}{2}, \tag{4.25}$$

where visual quality $\mathscr{Q}$ has a dynamic range of $[0,\ 1]$.

4.2 Results

We determined iteration and particle numbers experimentally. Fig. 4.5 shows the position and orientation errors under different iterations and particles. We chose particle and iteration numbers, $M = 20$ and $N = 8$, since they can balance the accuracy and computational time of our proposed method.

Table 4.1 summarizes the guidance accuracy of the five methods. The average accuracy or error of our proposed EPSO-based method (M5) was about 3.0 mm and 5.6° while the previous methods (M1, M2, M3, and M4) show at least 3.9 mm and 7.0°. Fig. 4.6 plots the guidance accuracy of different methods on Data 3. Beyond using GTDs for assessment, we further performed ten experiments where we did not generate ground truth for them. Fig. 4.7 shows the smoothness of the five methods evaluated on ten experiments. The average position and orientation smoothness of our method was 1.0 mm and 1.6°, which is significantly better than other methods M1 (4.0 mm and 2.1°), M2 (4.2 mm and 4.7°), M3 (3.7 mm and 3.8°), and M4 (2.0 mm and 2.6°). Fig. 4.8 (a) compares the visual quality of the five methods on ten experiments. The average quality of methods of M1, M2, M3, M4, and M5 were 0.15, 0.17, 0.23, 0.25, and 0.29, respectively. Our proposed method shows better visual quality than others. Fig. 4.9 further demonstrates that our method can obtain much better visual quality since the virtual images from our proposed approach resemble video images greatly better than other methods.

Additionally, Fig. 4.8 (b) displays the computational times of the five different methods. The current runtime of our proposed method was 0.71 seconds per frame (spf). The methods of M2, M3, and M4 were 0.45, 1.21, and 0.78 spf. The method of M1 can meet the real-time requirement during the endoscopic guidance.

Table 4.1: Comparison of guidance accuracy (position, orientation) in terms of the ground truth datasets and the estimated results of using the five different methods of M1 [4], M2 [5], M3 [8], M4 [9], and M5 (our EPSO-based method).

Methods	Data 1	Data 2	Data 3	Data 4	Data 5	**Average**
M1	(4.2 mm, 6.7°)	(5.3 mm, 8.8°)	(5.6 mm, 7.9°)	(6.0 mm, 9.6°)	(7.2 mm, 13.5°)	**(5.7 mm, 9.3°)**
M2	(3.8 mm, 6.1°)	(4.9 mm, 7.6°)	(5.4 mm, 6.8°)	(5.8 mm, 8.8°)	(6.8 mm, 12.9°)	**(5.3 mm, 8.4°)**
M3	(3.1 mm, 4.8°)	(3.9 mm, 5.8°)	(4.1 mm, 6.2°)	(4.6 mm, 9.5°)	(5.6 mm, 12.9°)	**(4.3 mm, 7.8°)**
M4	(2.4 mm, 4.3°)	(3.4 mm, 5.3°)	(4.2 mm, 5.6°)	(4.4 mm, 8.2°)	(5.1 mm, 11.5°)	**(3.9 mm, 7.0°)**
M5	(2.1 mm, 3.2°)	(2.8 mm, 3.8°)	(3.0 mm, 4.1°)	(3.2 mm, 7.9°)	(4.1 mm, 9.1°)	**(3.0 mm, 5.6°)**

4.3 Discussion

4.3.1 Effectiveness

Basically, the objective of this work is to effectively and accurately fuse multimodal information of pre-operative images, endoscopic video sequences, and positional sensor measurements to determine endoscope position and orientation during endoscopic surgery. Our proposed EPSO algorithm provides a more accurate and robust framework to guide the endoscope than previous approaches. We attribute such an advantageous performance of EPSO to several aspects.

First, we believe that EPSO is partly an association of PSO iterations and sequential Monte Carlo (SMC) sampling procedures, and hence it outperforms SMC sampling algorithms in endoscope navigation [8]. During SMC sampling procedures, a successful particle sampling depends heavily on the proposal distribution function [50]. Particles with large fitnesses or weights located in the useful area of the proposal distribution are usually sampled. In fact, the proposal distribution is suggested to be the dynamic transition distribution, which may incur particles with larger fitness that are not sampled when the useful area of the transition distribution stays at the tail of the observation distribution [50]. However, PSO performs more like a hierarchical sampling strategy which propagates the particles integrated with the newest observations [9], possibly resolving the particle impoverishment problem. Next, automatically or adaptively controlling evolutionary parameters is greatly helpful to update particles in iterations. The two acceleration factors, which were calculated based on the fitness value from the image intensity information, are more reasonable than setting them to 2.0 in standard PSOs [44]. Moreover, the inertia weight is also adaptively determined by both spatial distance constraint and image intensity information, resulting in more flexibly balancing the global and local search abilities and providing a reasonable velocity limitation to move particles. Finally, without any *resampling process* in our method, compared to SMC sampling or particle filtering, it is helpful to reduce the runtime of our method.

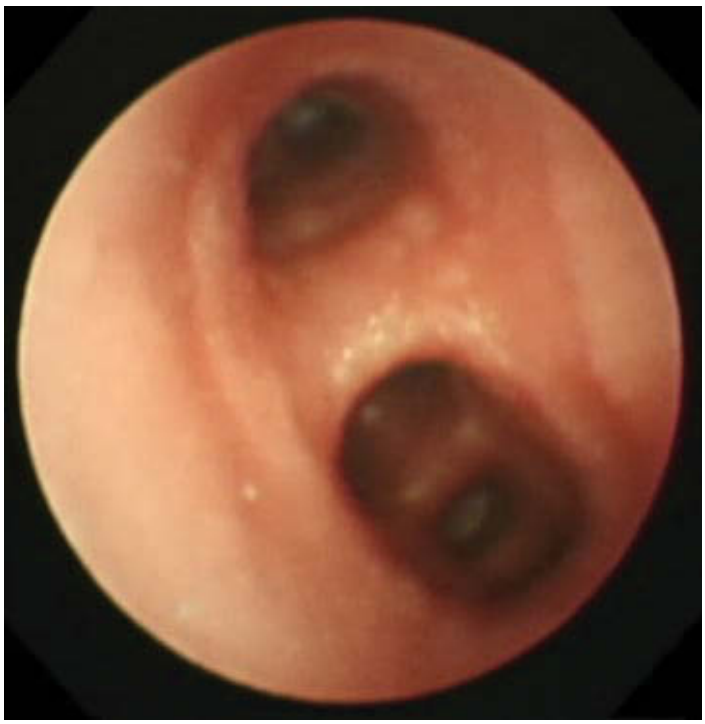
4.3.2 Potential Limitations

We must clarify the potential limitations of our proposed method. The major limitation is that the particle fitness might be incorrectly computed and evaluated during particle propagation since it depends somewhat on the quality of endoscopic camera images. However, image artifacts, e.g., complex reflectance and motion blurring, which possibly occur in endoscopic video sequences, might collapse the correct computation of the fitness value. On the other hand, the particle fitness value also indicates the particle's exploitation capacity that means

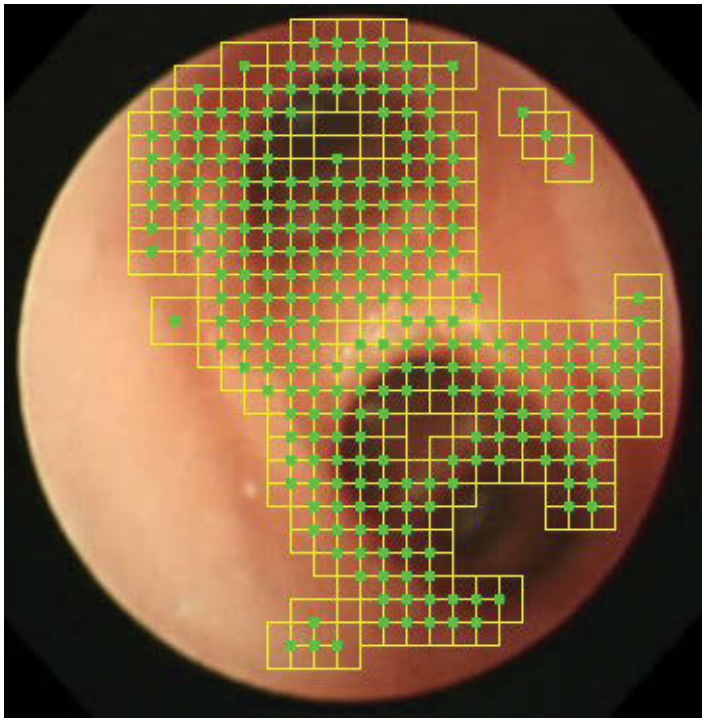
sufficient fitness values to obtain even better solutions from good particles: the higher the particle fitness, the more powerful particle exploitation. To tackle the incorrect fitness computation problem and obtain the powerful exploitation capacity, a more robust image similarity measure, which should be greatly insensitive to illumination changes or other image artifacts, is expected to develop in the future. Additionally, another open issue, which also remains very challenging, is the computational efficiency of our proposed method that is currently about 0.71 seconds per frame. In our proposed EPSO approach, the main computational effort lies in the calculation of each particle's fitness in the population, which is vastly time-consuming since 2-D virtual rendering images need to be generated by performing the volume rendering procedure. Note that we did neither speed optimizations nor multi-threading in our implementation. Furthermore, it is also possible to use graphics processing unit (GPU) techniques to accelerate our implementation to a real-time processing.

4.4 Conclusions

This article proposed an enhanced particle swarm optimization algorithm. It combines the current observation into the population propagation and refreshes its evolutionary parameters on the basis of the spatially distributed constraint of particles and the observation information during iterations. Therefore, it is more accurate and robust to approximate the optimal solution for a stochastic problem. With its application to the multimodal information fusion for the endoscopic guidance, the experimental results demonstrate that our proposed method provides a more advantageous guidance performance than state-of-the-art methods. The average position and orientation errors were greatly reduced from (4.3, 7.8) to (3.0 mm, 5.6°). The average position and orientation smoothnesses were also significantly improved from (3.7, 3.8) to (1.0 mm and 1.6°), which makes our guidance less jitter or jumps during endoscopic surgery. Future work includes improvement of the particle fitness calculation to correctly evaluate particles and reduction of the computational time of our proposed method.

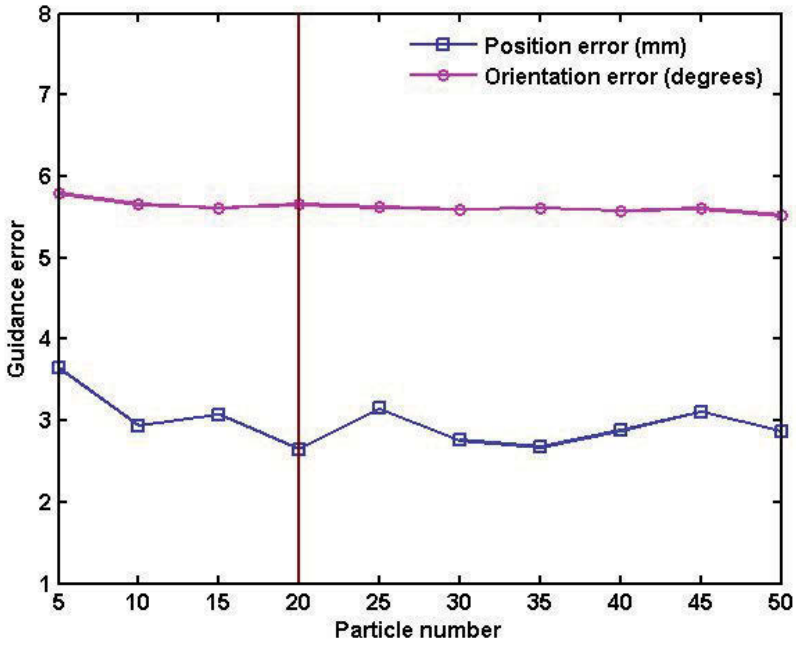


(a) Input camera image

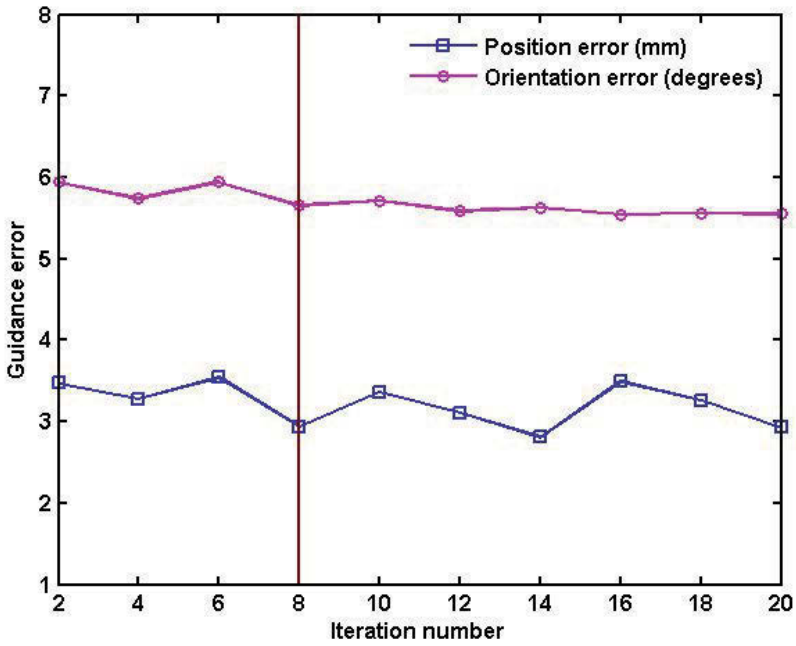


(b) Detected image patches

Figure 4.4: Detect image patches with specific structures from camera image $\mathbf{I}_k$ for the similarity or the fitness computation. A *yellow* square denotes a selected patch with its center (*green* dot).

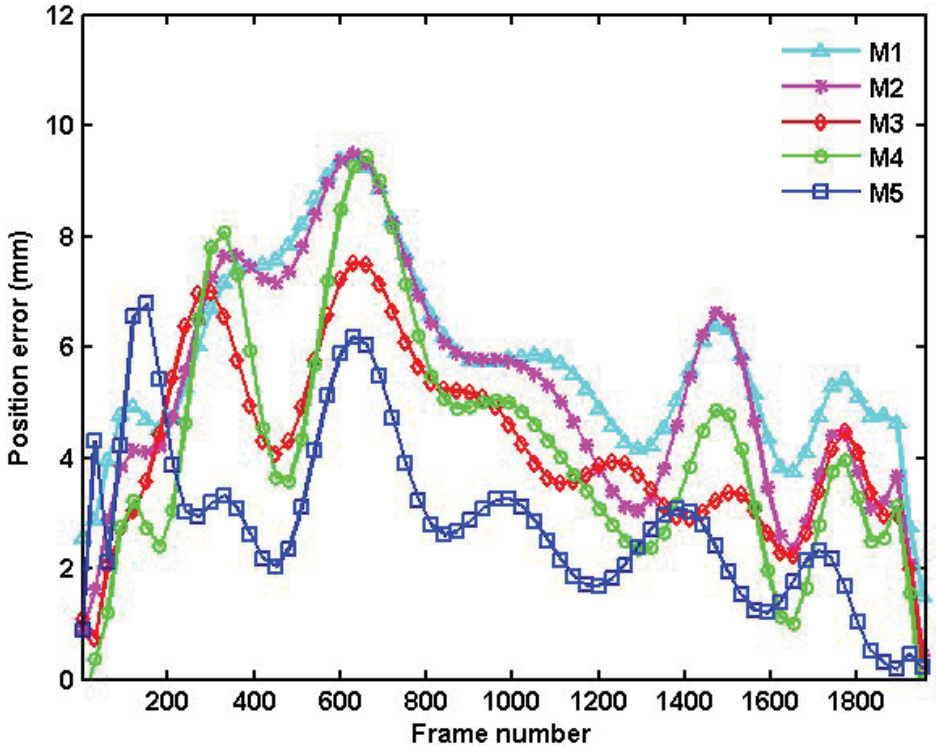


(a) Particles

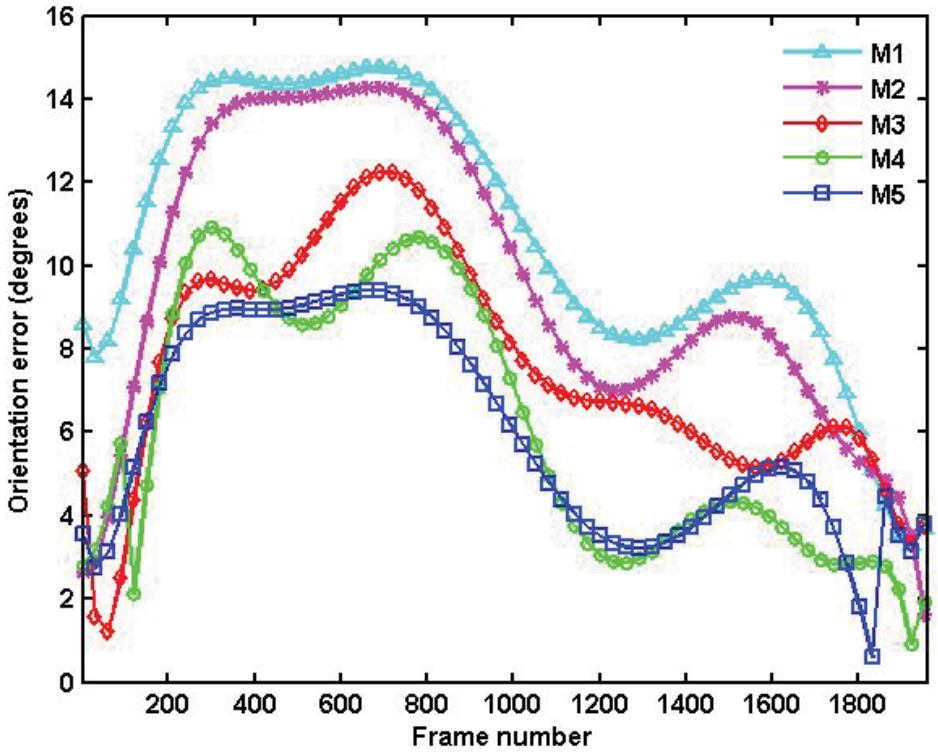


(b) Iterations

Figure 4.5: Experimentally determined particle and iteration numbers: $M = 20$ and $N = 8$.

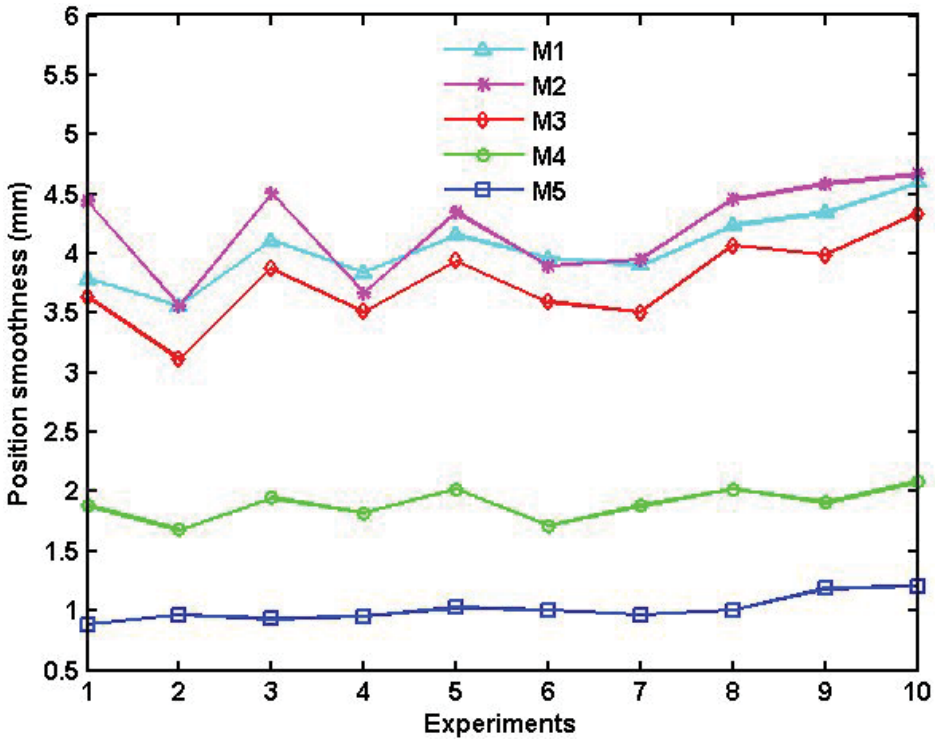


(a) Position error

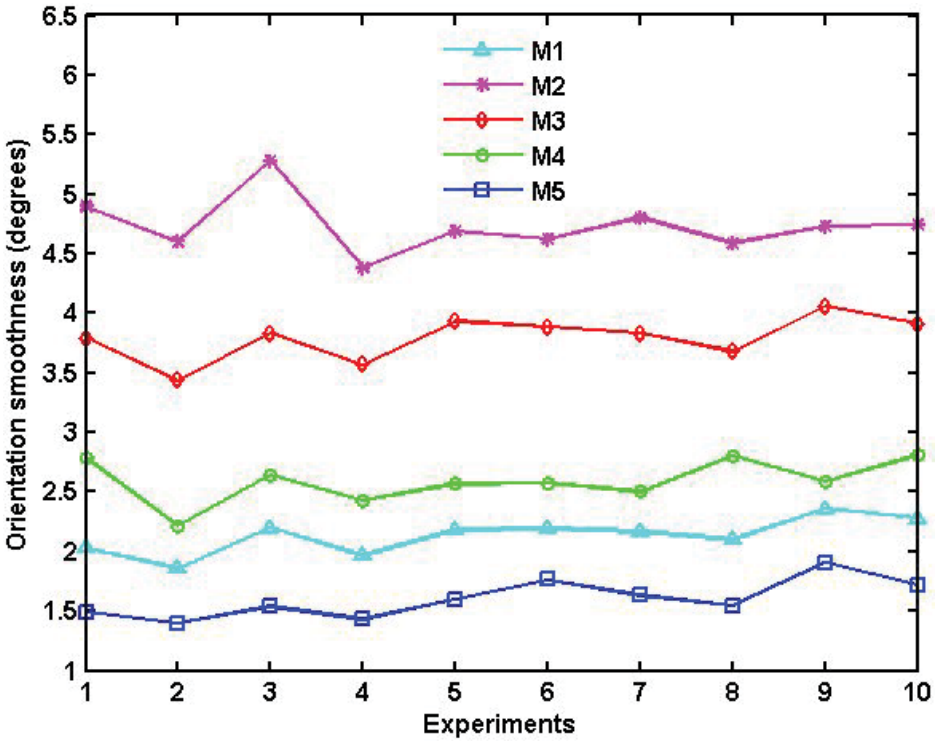


(b) Orientation error

Figure 4.6: Plotted the guidance accuracy of the five methods on Data 3.

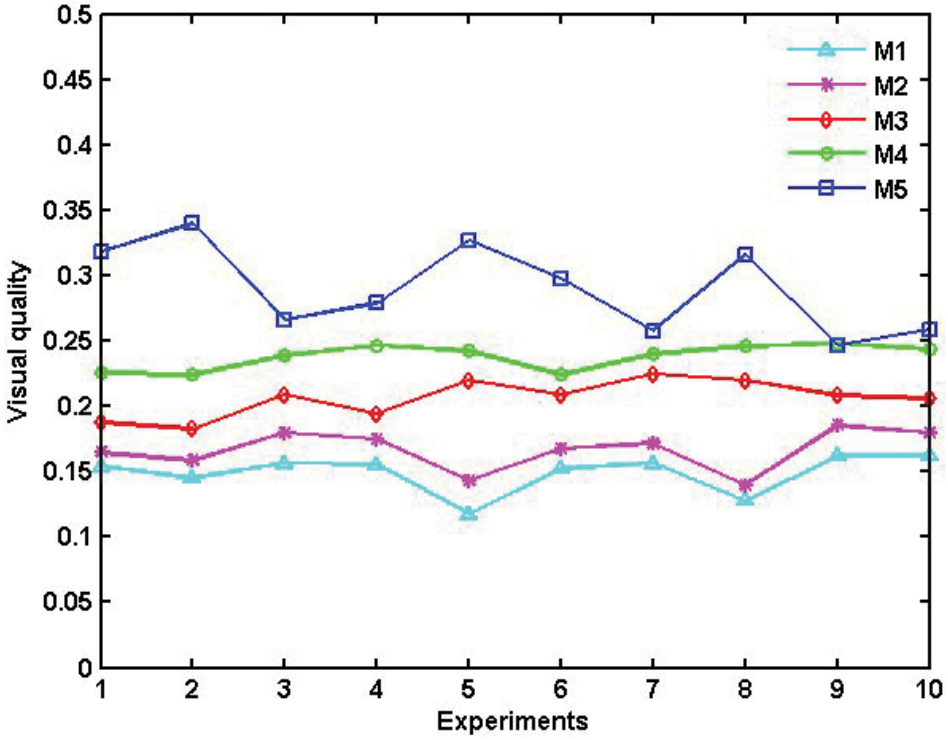


(a) Position smoothness

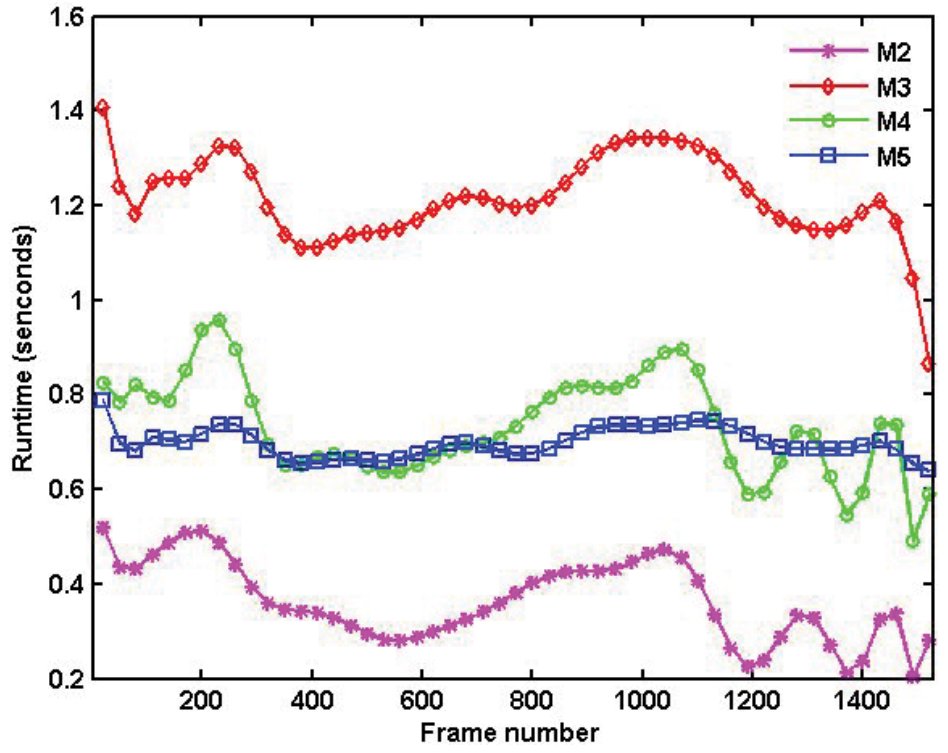


(b) Orientation smoothness

Figure 4.7: Position and orientation smoothness of the five methods on ten experiments.



(a) Visual quality



(b) Runtime

Figure 4.8: Comparison of visual quality and computational time of the five methods.

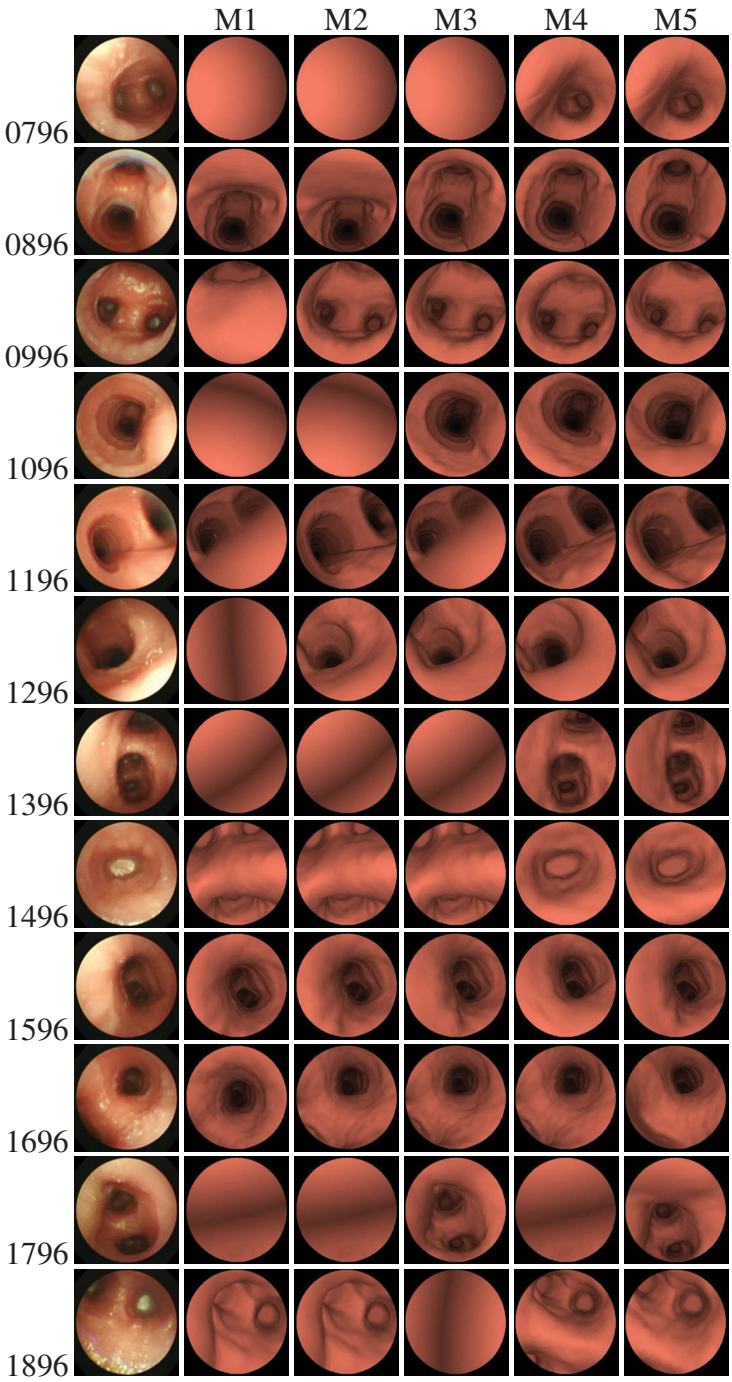


Figure 4.9: Visualized video and virtual images to investigate the visual quality of different methods. First column shows frame numbers selected uniformly at every 100 frames, and second column shows their corresponding real images. Other columns display virtual images generated from tracking results using the methods discussed above. Our method displays much better quality.

Chapter 5

Observation-driven differential evolution

This chapter 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.which corresponds to our journal paper titled “Observation-driven adaptive differential evolution and its application to accurate and smooth bronchoscope three-dimensional motion tracking” that was published in Medical Image Analysis.

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.

5.1 Purpose

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 [22, 23]. Although BIRT works well, it relies heavily on the video image quality that might be degenerated by image artifacts (e.g., motion 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 [4]. 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. Even though BEMT has been commercialized for clinical applications [4], 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.

As one of many evolutionary algorithms, differential evolution (DE) was originally developed by Storn and Price [51]. Nowadays it is still an active topic discussed for various optimization scenarios. A survey of DE algorithms and their practical applications has been published recently [52]. Due to its powerful and efficient implementation, it has been widely applied as a successful optimization technique to address multidimensional complex problem in the literature [53–55]. However, the DE's performance depends heavily on the evolutionary parameters of the mutation factor and the crossover rate. We here 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.

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 direction of each individual in the population during evolution, to enhance the DE performance; (2) We modified a structural similarity measure to compute the individual's fitness robustly. Since bronchoscopic video images are usually observed with the structural information such as bifurcation and fold shapes, we extract these shapes to boost the fitness computation and accurately characterize the individual performance in OADE; (3) 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 leaded into OADE to achieve a more robust, accurate, and smoothing bronchoscope tracking method; and (4) Additionally, our OADE algorithm can be suitable to track other endoscopes, e.g., sinuscope and colonscope.

5.1.1 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. To integrate the multi-modal information for BEMT, the relationship of these coordinate systems is established by:

$$^{CT}\mathbf{T}_C^k = {}^{CT}\mathbf{T}_{EMT}\,{}^{EMT}\mathbf{T}_S^k\,{}^{S}\mathbf{T}_C, \tag{5.1}$$

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.

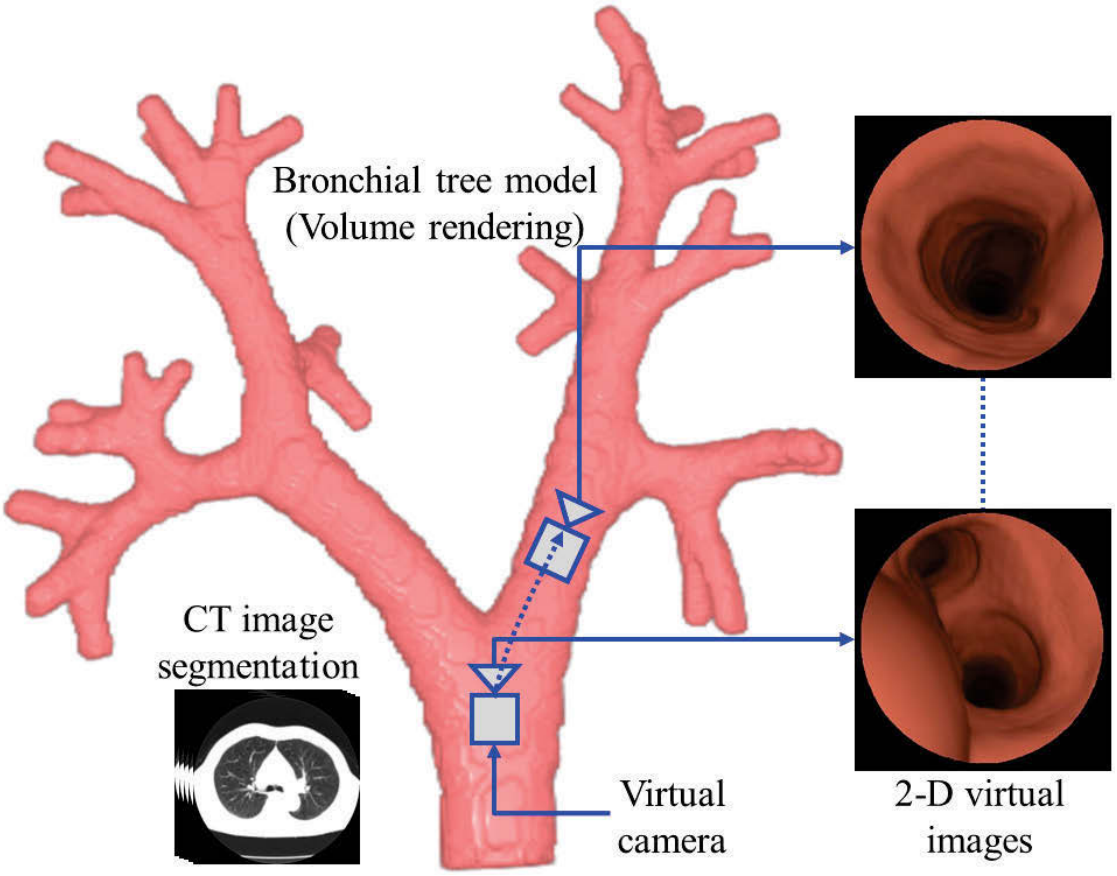


Figure 5.1: CT image segmentation for obtaining a 3-D bronchial tree model where a virtual camera flies through and generates 2-D virtual images by different camera poses.

${}^{EMT}\mathbf{T}_S^k$ can be represented by a homogeneous matrix:

$$ {}^{EMT}\mathbf{T}_S^k = \begin{pmatrix} {}^{EMT}\mathbf{R}_S^k & {}^{EMT}\mathbf{t}_S^k \\ \mathbf{0} & \mathbf{1} \end{pmatrix}_{4\times 4}, \tag{5.2} $$

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:

$$ {}^{CT}\mathbf{T}_C^k = \begin{pmatrix} {}^{CT}\mathbf{R}_C^k & {}^{CT}\mathbf{t}_C^k \\ \mathbf{0} & \mathbf{1} \end{pmatrix}_{4\times 4}, \tag{5.3} $$

where ${}^{CT}\mathbf{t}_C^k = [{}^{CT}t_C^x, {}^{CT}t_C^y, {}^{CT}t_C^z]_{3\times 1}$, 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 [56]. The quaternion is represented by a four-element vector, which has advantages of the compactness and the avoidance of discontinuous jumps, compared to other representations (e.g., Euler angles). The vector representation also provides efficient calculation by replacing matrices involved in the mutation and crossover steps in our OADE algorithm. Quaternion ${}^{CT}\mathbf{Q}_C^k$ to represent rotation ${}^{CT}\mathbf{R}_C^k$ is usually formulated as [56]:

$$ {}^{CT}\mathbf{R}_C^k \longleftrightarrow {}^{CT}\mathbf{Q}_C^k = [Q_0, Q_x, Q_y, Q_z]_{4\times 1}, \tag{5.4} $$

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 ${}^{CT}\mathbf{Q}_C^k$:

$$ {}^{CT}\mathbf{T}_C^k \longleftrightarrow \mathbf{X}^k = [\,{}^{CT}\mathbf{t}_C^k; \quad {}^{CT}\mathbf{Q}_C^k\,]_{7\times 1}. \tag{5.5} $$

Note that although the bronchoscope motion is parametrized by a 7-D vector, it still contains 6DoF movement information.

5.1.2 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 [22] (Fig. 5.1). 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.

5.2 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 state-of-the-art DE algorithms and show an idea to improve DEs.

Basically, the DE algorithm propagates a population of individuals or vectors $\{\mathbf{X}_{i,j}|\mathbf{X}_{i,j} \in \Re^N\}_{i=1}^P$ (P is the population size, j is the generation index, $j = 1,\cdots,G$, and N is the dimension of the vector) toward the global optimum during any stochastic optimization procedures. After initializing population $\{\mathbf{X}_{i,j}|\mathbf{X}_{i,j} \in \Re^N\}_{i=1}^P$, 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. 5.5), we replace $\mathbf{X}_{i,j}$ with $\mathbf{X}_{i,j}^k$ and obtain population $\mathscr{H}_j^k$=$\{\mathbf{X}_{i,j}^k|\mathbf{X}_{i,j}^k \in \Re^N\}_{i=1}^P$.

5.2.1 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 [52]:

$$\mathbf{V}_{i,j}^k = \mathbf{X}_{r_i^1,j}^k + F_i\left(\mathbf{X}_{r_i^2,j}^k - \mathbf{X}_{r_i^3,j}^k\right) \tag{5.6}$$

$$\mathbf{V}_{i,j}^k = \mathbf{X}_{best,j}^k + F_i\left(\mathbf{X}_{r_i^1,j}^k - \mathbf{X}_{r_i^2,j}^k\right), \tag{5.7}$$

$$\mathbf{V}_{i,j}^k = \mathbf{X}_{i,j}^k + F_i\left(\mathbf{X}_{best,j}^k - \mathbf{X}_{i,j}^k\right) + F_i\left(\mathbf{X}_{r_i^1,j}^k - \mathbf{X}_{r_i^2,j}^k\right), \tag{5.8}$$

$$\mathbf{V}_{i,j}^k = \mathbf{X}_{best,j}^k + F_i\left(\mathbf{X}_{r_i^1,j}^k - \mathbf{X}_{r_i^2,j}^k\right) + F_i\left(\mathbf{X}_{r_i^3,j}^k - \mathbf{X}_{r_i^4,j}^k\right), \tag{5.9}$$

$$\mathbf{V}_{i,j}^k = \mathbf{X}_{r_i^1,j}^k + F_i\left(\mathbf{X}_{r_i^2,j}^k - \mathbf{X}_{r_i^3,G}^k\right) + F_i\left(\mathbf{X}_{r_i^4,j}^k - \mathbf{X}_{r_i^5,j}^k\right). \tag{5.10}$$

Eqs. 5.6∼5.10 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 $r_i^1, r_i^2, r_i^3, r_i^4$, and r_i^5 are mutually exclusive integers chosen randomly from set $\{1,\cdots,i-1,i+1,\cdots,P\}$. The part in brackets in Eqs. 5.6∼5.10, e.g., $(\mathbf{X}_{r_i^1,j}^k - \mathbf{X}_{r_i^2,j}^k)$, represents the difference vector; $\mathbf{X}_{best,j}^k$ is the best individual at generation j, and F_i is the mutation factor.

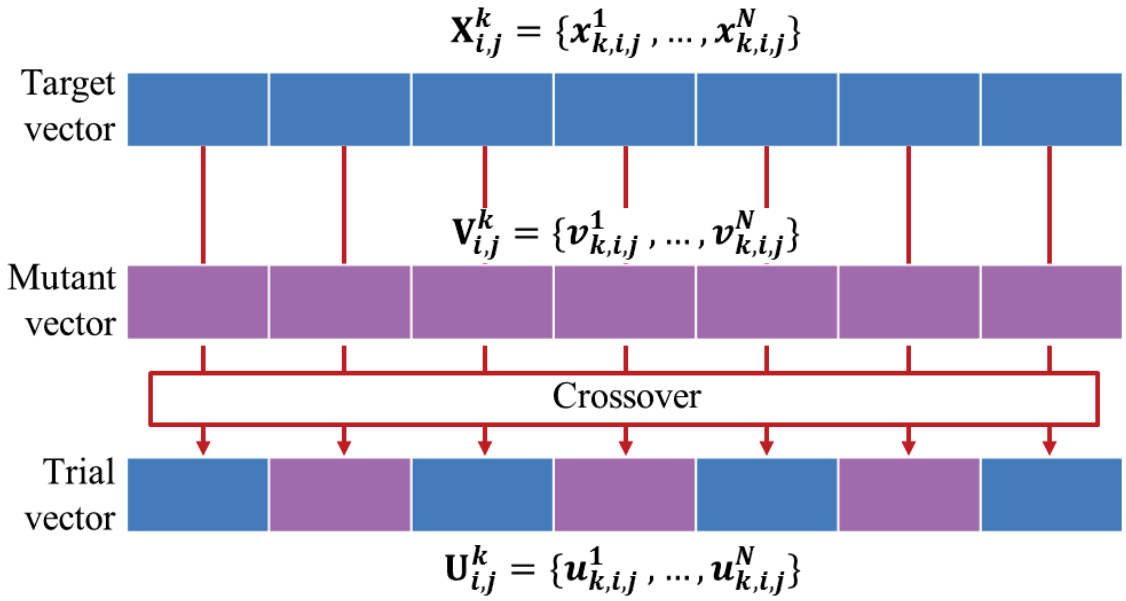


Figure 5.2: Illustration of crossover step in DE.

5.2.2 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 by exchanging the N-dimensional information between the target and mutant vectors (Fig. 5.2). The existing DE algorithms use either the exponential or binomial crossover [52]. In this work, we use the binomial crossover and perform it to generate trial vector $\mathbf{U}_{i,j}^{k} = \{u_{k,i,j}^{1}, \cdots, u_{k,i,j}^{N}\}$ in terms of target vector $\mathbf{X}_{i,j}^{k} = \{x_{k,i,j}^{1}, \cdots, x_{k,i,j}^{N}\}$ and mutant vector $\mathbf{V}_{k,i,j}^{k} = \{v_{k,i,j}^{1}, \cdots, v_{k,i,j}^{N}\}$:

$$u_{k,i,j}^{d} = \begin{cases} v_{k,i,j}^{d} & \text{if } (ran_d[0,\ 1] \leq C_r) \text{ or } (d = d_{ran}) \\ x_{k,i,j}^{d} & \text{otherwise} \end{cases}, \tag{5.11}$$

where $ran_d[0,\ 1]$ 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_{ran} is randomly selected from set $\{1, 2, \cdots, d, \cdots, N\}$.

5.2.3 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.

Algorithm 1: Standard DE algorithm

0. Set population size P;
1. Control parameters: $F_i = constant$, $C_r = constant$;
2. At generation $j = 0$, initialize $\mathscr{H}_0^k = \{\mathbf{X}_{i,0}^k | \mathbf{X}_{i,0}^k \in \Re^N\}_{i=1}^P$;
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. 5.6∼5.10;

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. 5.11;

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. 5.12;

end
$j = j + 1$;
end while

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:

$$\mathbf{X}_{i,j+1}^k = \begin{cases} \mathbf{U}_{i,j}^k & \text{if } W(\mathbf{U}_{i,j}^k, \mathbf{I}^k) \geq W(\mathbf{X}_{i,j}^k, \mathbf{I}^k) \\ \mathbf{X}_{i,j}^k & \text{otherwise} \end{cases}, \tag{5.12}$$

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 [51, 52].

5.2.4 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 [52]. Several adaptive differential evolution (ADE) methods were proposed and obtained better performance [57–59, 53, 7]. 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. [60], the optimal solution space of any stochastic optimization problems should be integrated into the current observation. Unfortunately, no state-of-the-art 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 getting trapped 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 5.3.

5.3 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.

5.3.1 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 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

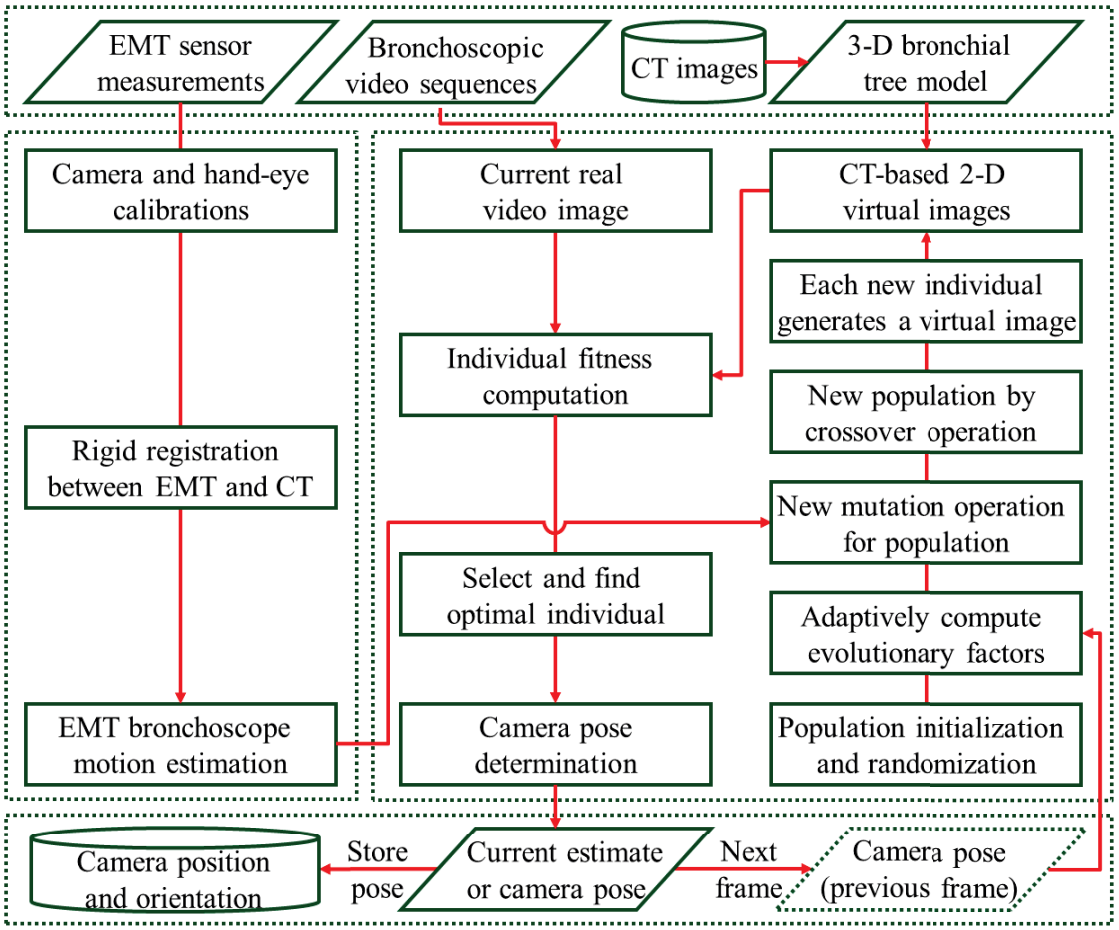


Figure 5.3: Processing flowchart of our OADE tracking method.

3-D motion pose. Fig. 5.3 illustrates the processing flowchart of our approach with its different steps, which are discussed in the subsequent sections.

5.3.2 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. 5.1, 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 ${}^{S}\mathbf{T}_{C}$ [6] and use a fiducial-based method to implement a rigid registration to calculate ${}^{CT}\mathbf{T}_{EMT}$ [4]. 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.

5.3.3 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. 5.5 and obtain $\mathscr{H}_j^1$ at generation j:

$$\mathscr{H}_j^1 = \{{}^{CT}\mathbf{T}_C^1 \longleftrightarrow \mathbf{X}_{i,j}^1\}_{i=1}^P. \tag{5.13}$$

Note that during any stochastic optimization problem, the diversity of the population in the DE methods plays a positive role in the optimization performance. To enhance the population diversity, we randomize population $\mathscr{H}_j^1$ in terms of the normal distribution and achieve population $\mathscr{H}_j^k$ at frame k:

$$\mathscr{H}_j^k = \{\mathbf{X}_{i,j}^k\}_{i=1}^P, \quad \mathbf{X}_{i,j}^k = \Gamma(\mathbf{X}_{i,j}^1, \pi_i\Delta), \tag{5.14}$$

where π_i is a normally distributed random number, Δ is a predefined vector, and Γ 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 $k=1$.

5.3.4 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 [61, 57, 62, 58, 59, 53, 7]. Many mutation strategies were proposed, as shown in Eqs. 5.6∼5.10. The advantages and limitations of different mutation schemes were discussed [52]. In our framework, we modify the strategy of $DE/target-to-best/1$ (Eq. 5.8) that has good convergence performance due to its usage of the best solution or individual information [62]. 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 $DE/target-to-best/1$:

$$\begin{cases} \mathbf{V}_i^b = F_i^b\left(\mathbf{X}_{best,j}^k - \mathbf{X}_{i,j}^k\right) \\ \mathbf{V}_i^r = F_i^r\left(\mathbf{X}_{r_i^1,j}^k - \mathbf{X}_{r_i^2,j}^k\right) \\ \mathbf{V}_{i,j}^k = \mathbf{X}_{i,j}^k + \underbrace{\Omega_i\left(\mathbf{E}^k - \mathbf{E}^{k-1}\right)}_{\text{Sensor observation}} + \mathbf{V}_i^b + \mathbf{V}_i^r \end{cases}, \tag{5.15}$$

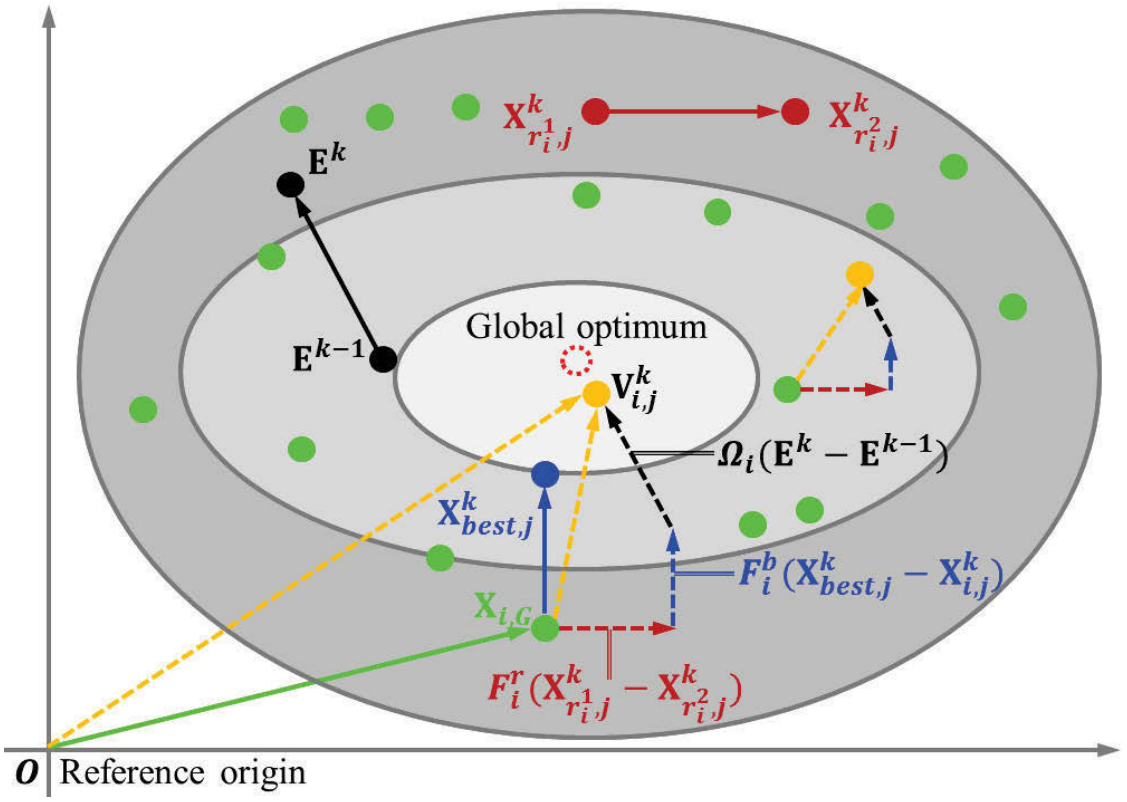


Figure 5.4: Illustration of our mutation strategy to generate mutant vector $\mathbf{V}^k_{i,j}$ in accordance with three perturbations including observation $\Omega_i\left(\mathbf{E}^k - \mathbf{E}^{k-1}\right)$ and two difference vectors of $F_i^b\left(\mathbf{X}^k_{best,j} - \mathbf{X}^k_{i,j}\right)$ and $F_i^r\left(\mathbf{X}^k_{r_i^1,j} - \mathbf{X}^k_{r_i^2,j}\right)$.

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 Ω_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 $\mathbf{X}^k_{best,j}$, $\mathbf{X}^k_{i,j}$ and current camera observation (image) $\mathbf{I}^k$:

$$F_i^b = \frac{2W\left(\mathbf{X}^k_{best,j}, \overbrace{\mathbf{I}^k}^{\text{Camera observation}}\right)}{\left(W\left(\mathbf{X}^k_{best,j}, \mathbf{I}^k\right) + W\left(\mathbf{X}^k_{i,j}, \mathbf{I}^k\right)\right)}, \tag{5.16}$$

$$F_i^r = \frac{2W\left(\mathbf{X}^k_{i,j}, \mathbf{I}^k\right)}{\left(W\left(\mathbf{X}^k_{best,j}, \mathbf{I}^k\right) + W\left(\mathbf{X}^k_{i,j}, \mathbf{I}^k\right)\right)}. \tag{5.17}$$

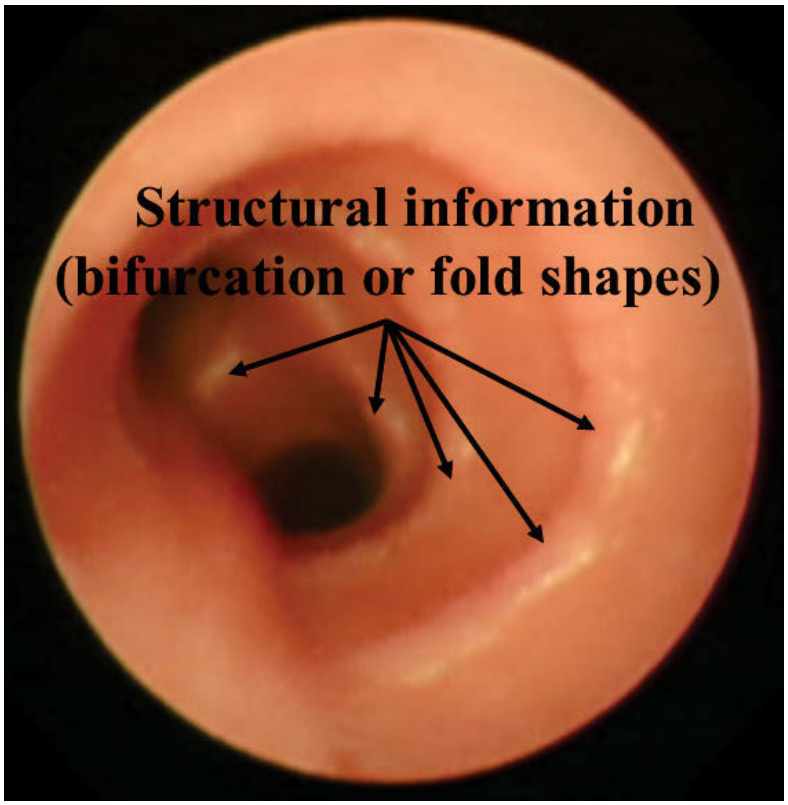


Figure 5.5: An example of one bronchoscopic image with structural information

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 Ω_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 $\Omega_i \left(\mathbf{E}^k - \mathbf{E}^{k-1}\right)$ 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 $(\mathbf{X}_{r_i^1,j}^k - \mathbf{X}_{r_i^2,j}^k)$. 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. 5.4 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 [53], we can calculate it adaptively by:

$$C_r = \frac{\left(W\left(\mathbf{X}_{i,j}^k, \mathbf{I}^k\right) + W\left(\mathbf{V}_{i,j}^k, \mathbf{I}^k\right)\right)}{2}, \tag{5.18}$$

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

5.3.5 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 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 $\mathbf{U}_{i,j}^k$):

$$W\left(\mathbf{X}_{i,j}^k, \mathbf{I}^k\right) = S\left(\mathbf{I}_v(\mathbf{X}_{i,j}^k), \mathbf{I}^k\right), \tag{5.19}$$

where $S(\cdot)$ is the intensity similarity function. We define $S(\cdot)$ using the structural similarity (SSIM) measure [63] and introduce a modified structural similarity function.

Based on our recent work [64], we compute intensity similarity $S(\cdot)$ relative to $\mathbf{X}_{i,j}^k$ and $\mathbf{I}^k$:

$$S\left(\mathbf{I}_v(\mathbf{X}_{i,j}^k), \mathbf{I}^k\right) = \frac{1}{\phi} \sum_{L_a^k \in \{L_a^k\}_{a=1}^{\phi}} \frac{1}{\left|L_a^k\right|} \sum_{L_a^k} M, \tag{5.20}$$

where ϕ is the number of selected image patches $\{L_a^k\}_{a=1}^{\phi}$ for calculating the local similarity and $|L_a^k|$ is the number of pixels in patch L_a^k. Patch similarity measure M is defined based on the SSIM measure, which was significantly better than a MSE-based measure [63]. We compute M by:

$$M = \frac{(2\xi_k\xi_v + C_1)\left(2\delta_{k,v} + C_2\right)}{\left(\xi_k^2 + \xi_v^2 + C_1\right)\left(\delta_k^2 + \delta_v^2 + C_2\right)}, \tag{5.21}$$

where $\delta_{k,v}$ is the correlation between $\mathbf{I}^k$ and $\mathbf{I}_v(\mathbf{X}_{i,j}^k)$ at patch L_a^k; ξ_k and ξ_v are the mean intensity values, δ_k and δ_v are the intensity variances, and C_1 and C_2 are constants. Fig. 5.6 shows an example of structural patch extraction and selection from Fig. 5.5.

In general, Eqs. 5.20 and 5.21 shows a robust 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.

5.3.6 Camera Pose Determination

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

$$\mathbf{X}^k_{*,*}=\arg\max_{\mathbf{X}^k_{*,j+1}\in\mathscr{H}^k} W(\mathbf{X}^k_{*,j+1},\mathbf{I}^k). \tag{5.22}$$

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

$$\mathbf{X}^k_{*,*}\longmapsto {}^{CT}\tilde{\mathbf{T}}^k_C=\begin{pmatrix} {}^{CT}\tilde{\mathbf{R}}^k_C & {}^{CT}\tilde{\mathbf{t}}^k_C \\ \mathbf{0} & \mathbf{1}\end{pmatrix}. \tag{5.23}$$

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 and We set the termination condition as G generations. In OADE, it must render 2-D virtual image $\mathbf{I}_v(\mathbf{X}^k_{i,j})$ and compute fitness $S\left(\mathbf{I}_v(\mathbf{X}^k_{i,j+1}),\mathbf{I}^k\right)$ for each individual in population $\{\mathbf{X}^k_{i,j+1}\}^P_{i=1}$ 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, $W(\mathbf{X}^k_{*,j+1},\mathbf{I}^k)$ and $W(\mathbf{X}^k_{*,j+2},\mathbf{I}^k)$. If $W(\mathbf{X}^k_{*,j+1},\mathbf{I}^k)=W(\mathbf{X}^k_{*,j+2},\mathbf{I}^k)$, 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**.

5.4 Experiments

A dynamic phantom that can simulate respiratory motion was constructed to evaluate our proposed method [3]. A bronchoscope (BF-P260F, Olympus, Tokyo) was used to acquire 362×370 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. [4], directly using absolute EMT sensor outputs; (2) Mori et al. [5], integrating absolute EMT sensor outputs into image registration; (3) Luo et al. [6], combining either absolute or relative (inter-frame) EMT sensor outputs with image registration; (4) Luo et al. [3], utilizing sequential Monte Carlo (SMC) methods to fuse inter-frame EMT sensor outputs with video image; (5) Zhang et al. [7], a method of adaptive differential evolution with optional external archive; (6) our method, as discussed in Section 5.3. Note that we compared SMC [3] and our proposed to show different optimization frameworks for bronchoscope tracking. We also compared our modified similarity function (Eq. 5.20) and a modified mean squared error (MoMSE) measure [24] 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 [7].

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 [3]. 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 (τ, θ) by:

$$\tau = \|\mathbf{t} - \mathbf{t}_G\|, \ \theta = \arccos((trace(\mathbf{R}\mathbf{R}_G{}^T) - 1)/2), \tag{5.24}$$

where τ is the Euclidean distance between ground truth position $\mathbf{t}_G$ and estimate $\mathbf{t}$ and θ 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:

$$\mu = \frac{\sum_{k=1}^{K-1} \|\mathbf{t}_{k+1} - \mathbf{t}_k\|}{K-1}, \tag{5.25}$$

$$\psi = \frac{\sum_{k=1}^{K-1} \arccos((trace(\mathbf{R}_{k+1}{\mathbf{R}_k}^T) - 1)/2)}{K-1}, \tag{5.26}$$

where K is the frame number. Large μ and ψ 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 ${}^{CT}\tilde{\mathbf{T}}_C^k$ 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 γ on the basis of a universal image quality index [65]:

$$\gamma = \frac{1}{2}\left(1 + \frac{4\vartheta_{k,v}\varepsilon_k\varepsilon_v}{\left(\vartheta_k^2 + \vartheta_v^2\right)\left(\varepsilon_k^2 + \varepsilon_v^2\right)}\right), \tag{5.27}$$

where $\vartheta_{k,v}$ is the correlation between image $\mathbf{I}^k$ and its virtual image $\mathbf{I}_v$ corresponding to estimated pose ${}^{CT}\tilde{\mathbf{T}}_C^k$, ϑ_r and ϑ_v are the covariance, and ε_k and ε_v are the average values.

5.5 Results

In our OADE tracking method, we experimentally determine population size P and generation number G. Figs. 5.7 $\sim$ 5.8 show the tracking error and processing time results under different population sizes and generation numbers. The tracking errors of $P = 25$, 30, and 40 were 2.165, 2.189, and 2.084 mm, respectively. We used population size $P = 25$ for the computational efficiency. According to Fig. 5.8(a), the tracking errors of $G = 2$, 3, 8, and 9 were 2.797, 2.756, 2.543, and 2.474 mm, respectively. Although $G = 9$ can obtain the best tracking accuracy, we still initially set generation number $G = 3$ for reducing the computational time. As discussed above (Section 5.3.6), 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, $G = 2$. We run iterations at two generations and obtain two best individual fitness values from each generation. If the second best individual fitness value equals the first one, iterations at the third generation will not be performed, i.e., $G = 2$. Otherwise, the generation number is not adjusted, i.e., $G = 3$. Also note that various optimization scenarios usually require different the generation number and population size due to different dimensional spaces. We experimentally calculated P and G since there are no any gold standards to select them. Our choices of $P = 25$ and $G = 2$ or 3 generally balance the accuracy and runtime.

Fig. 5.9 displays the tracking errors of using the six approaches in Experiments (or Datasets) 5 and 20. Fig. 5.10 shows an example of the tracking smoothness of Experiment 6. Fig. 5.11 illustrates the tracking visual quality of using the six methods in Experiments 8 and 14. Table 5.1 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. 5.12 investigates different optimization frameworks and similarity measures for bronchoscope tracking. The OADE algorithm works better than SMC (Luo et al. [3]) 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. 5.13 compares the tracking smoothness of the six methods during 21 experiments. Fig. 5.14(a) shows the visual quality of each method in them. Fig. 5.14(b) investigates the computational times of each method. The processing times of the five methods of Mori et al. [5], Luo et al. [6], Luo et al. [3], and Zhang et al. [7] were about 0.69, 1.13, 1.83, 0.97, and 0.92 seconds per frame (spf), respectively. Compared to [66], 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. [4] is real-time processing.

Fig. 5.15 illustrates the estimated camera trajectories of the methods of Luo et al. [3] and our OADE framework on a pre-built 3-D bronchial tree model. Fig. 5.16 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. 5.16 were generated with black regions (in circles) since the virtual camera with its estimated pose was outside the bronchial tree volume. Both Figs. 5.15 and 5.16 further prove that our OADE method estimates bronchoscope camera movements significantly better than the other methods.

5.6 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 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.

5.6.1 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. The population diversity determines the exploration ability. 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 Figs. 5.17 $\sim$ 5.19.

Furthermore, for the state-of-the-art 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 helpful to maintain the population diversity. In contrast to our previous SMC-based tracking method [3], 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 (Figs. 5.17 $\sim$ 5.19).

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 [3] since the structural similarity measure was proved to significantly outperform MSE [63].

5.6.2 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. 5.20). 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 bronchoscope has been already navigated to the target regions at the correct airway, we could reduce the computational time by sacrificing tracking accuracy.

Additionally, we did not evaluate our method on patient data. Such clinical validation is undoubtedly optimal. Unfortunately, obtaining clinical data is really difficult since limited clinical data exist worldwide, even though a commercially available EMT-based bronchoscopy system, superDimension, has been increasingly introduced in operating rooms. Animal experiments might be another option to validate our proposed tracking framework in the future. We also point out that the current tracking error is relative error since we definitely introduce error in our ground truth datasets. Obtaining real ground truth data is very challenging.

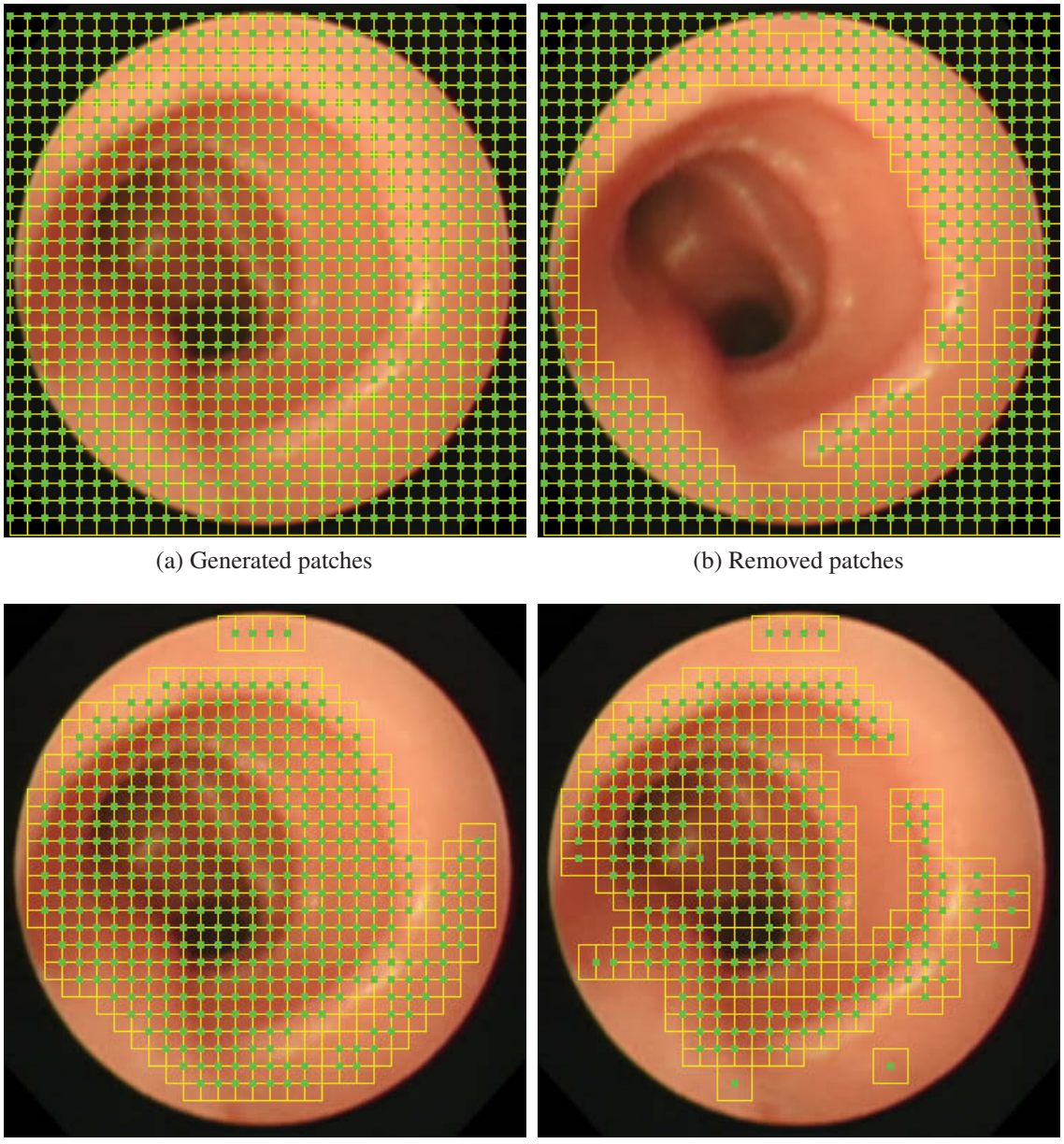
In summary, we proposed a new bronchoscope 3-D motion tracking scheme that bases on an observation-driven adaptive differential evolution algorithm that incorporates current sensor measurement and video image and can adaptively refresh its control parameters based on image intensity information during iterations. All experimental results demonstrates that our methods provides a more advantageous tracking performance than state-of-the-art methods.

Visualization quality assessment/evaluation: Two ideas to deal with the local minima problem: (1) The endoscope movements are usually considered as a reciprocate memorize and store motion parameters and images to recognize the same place the endoscope passed again, then we can reinitialize the tracking optimization and (2) manifold learning-based methods: recognize the bronchoscopic images and embedded into the subspace find the closest point. We may also develop an on-line visualization quality assessment mechanism: to remind physicians when the endoscope collide with the lumen walls

5.7 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 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.



(a) Generated patches (b) Removed patches

(c) Remaining patches (d) Selected patches

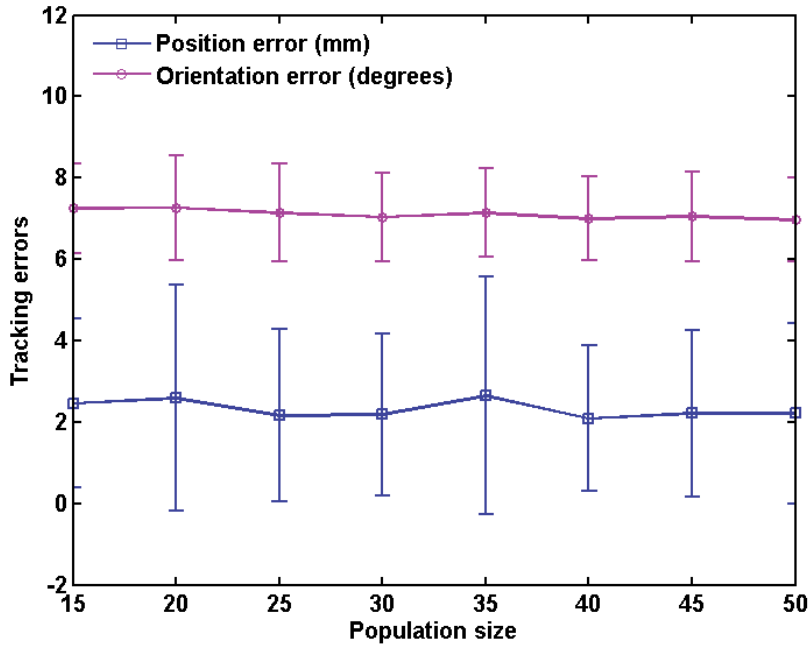
Figure 5.6: Extract structural regions in one bronchoscopic image to calculate similarity during fitness computation step. One *yellow* square denotes one patch and a *green* point is its center.

Algorithm 2: OADE-based bronchoscope tracking

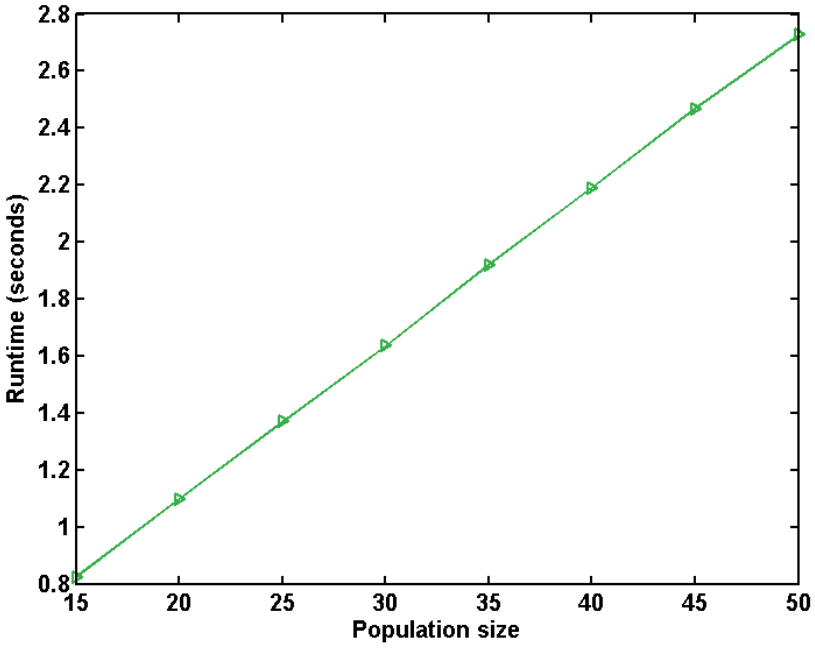
Input: Bronchoscopic video sequences, EMT sensor measurements, and CT images

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. 5.13∼5.14);
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. 5.19∼5.21), 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. 5.15∼5.17);
      ❹ Crossover: Get $\mathbf{U}_{i,j}^k$ (Eqs. 5.11 and 5.18);
      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. 5.19∼5.21) 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. 5.22∼5.23):
  Find current best solution $\mathbf{X}_{*,*}^k$ from $\mathscr{H}^k = \{\mathbf{X}_{*,j+1}^k\}_{j=1}^G$;
  $k = k + 1$;
end

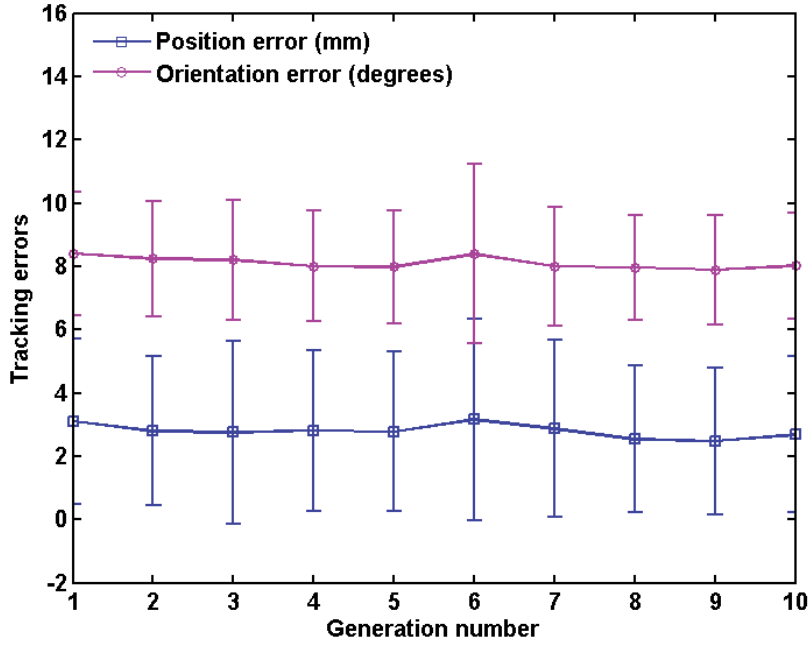


(a) Error of different population sizes

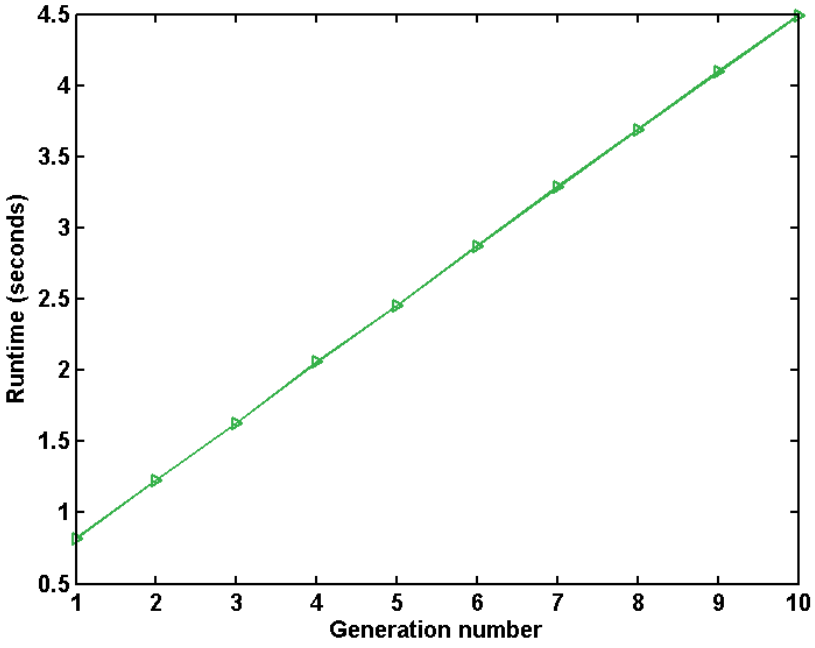


(b) Time of different population sizes

Figure 5.7: Experimentally determined population size: $P = 25$.



(a) Error of different generation numbers

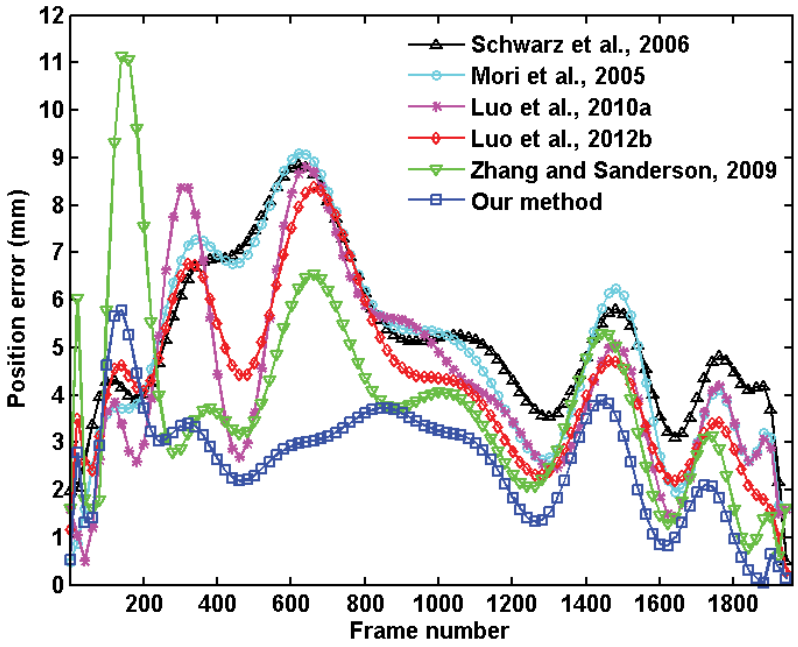


(b) Time of different generation numbers

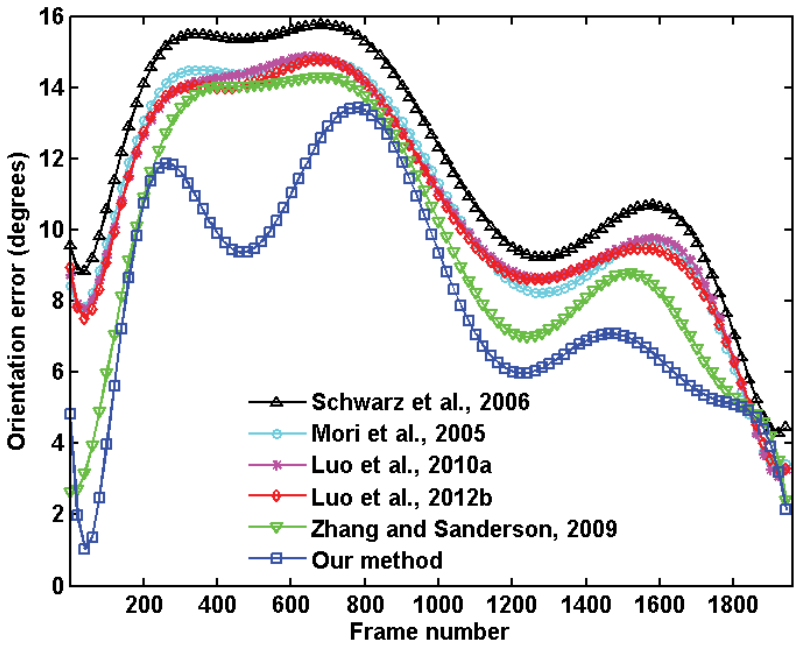
Figure 5.8: Experimentally determined generation number $G = 2$ or 3.

Table 5.1: 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. 5.24∼5.27. Our OADE outperformed other methods.

Compared approaches	Average accuracy		Average smoothness		Average visual quality γ
	Position τ (mm)	Orientation θ ($^\circ$)	Position μ (mm)	Orientation ψ ($^\circ$)	
[4]	5.10±2.65	11.46±8.83	4.67±0.90	3.57 ±0.39	0.636±0.008
[5]	4.43±3.29	10.98±9.45	4.80±0.80	4.91±0.35	0.668±0.086
[6]	4.39±4.23	10.63±9.94	4.27±0.86	4.29±0.36	0.680±0.094
[3]	3.96±3.60	10.47±8.82	4.08±0.59	2.65±0.33	0.707±0.095
[7]	3.63±3.21	9.53±7.55	1.93±0.58	2.35±0.96	0.715±0.004
Our method	2.89±2.87	8.18±6.29	1.62±0.53	2.11±0.69	0.741±0.005

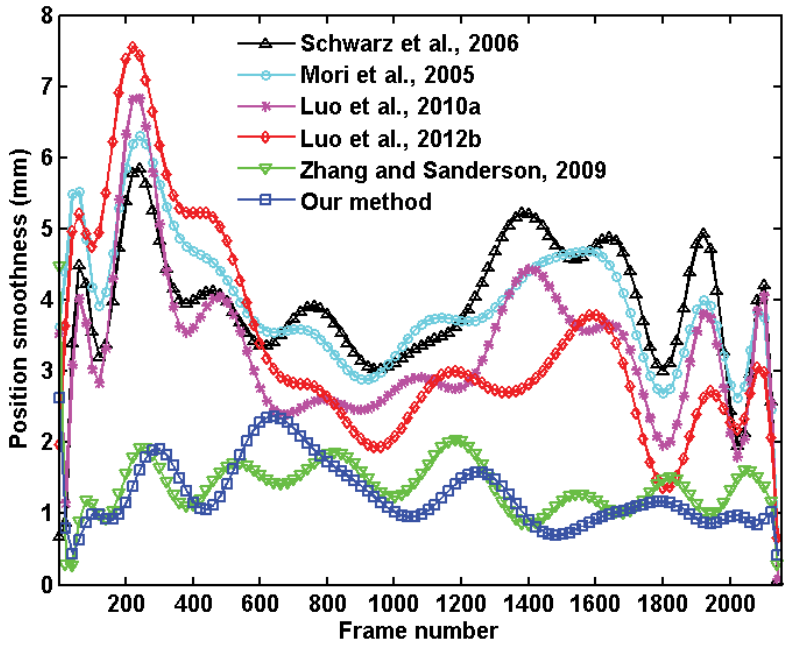


(a) Position error of Experiment 20

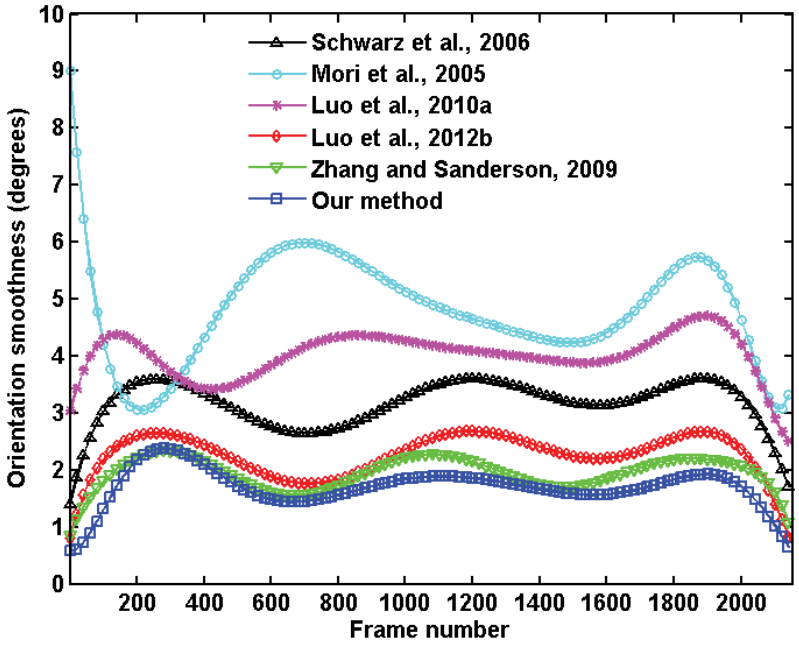


(b) Orientation error of Experiment 20

Figure 5.9: Plotted tracking error of six methods evaluated in Experiment 20.

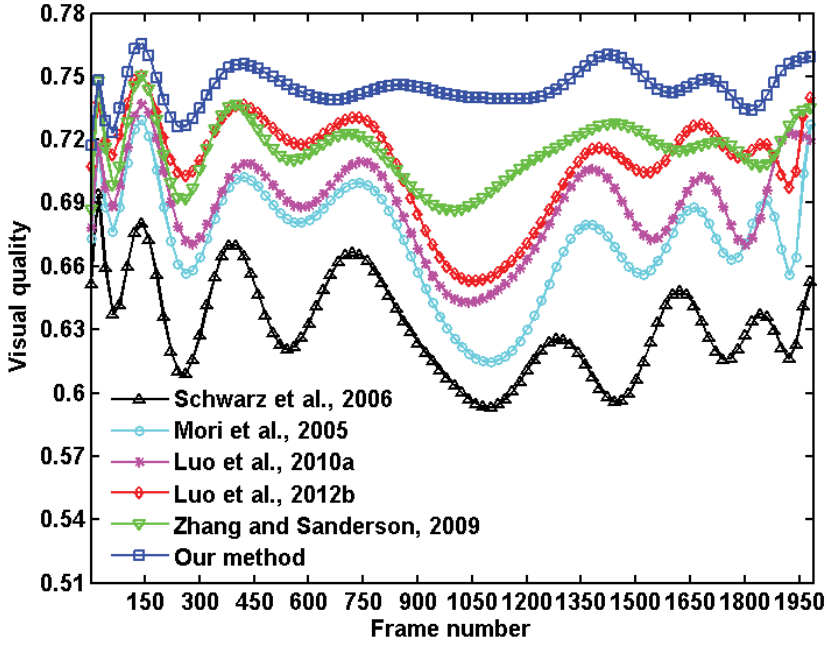


(a) Position smoothness

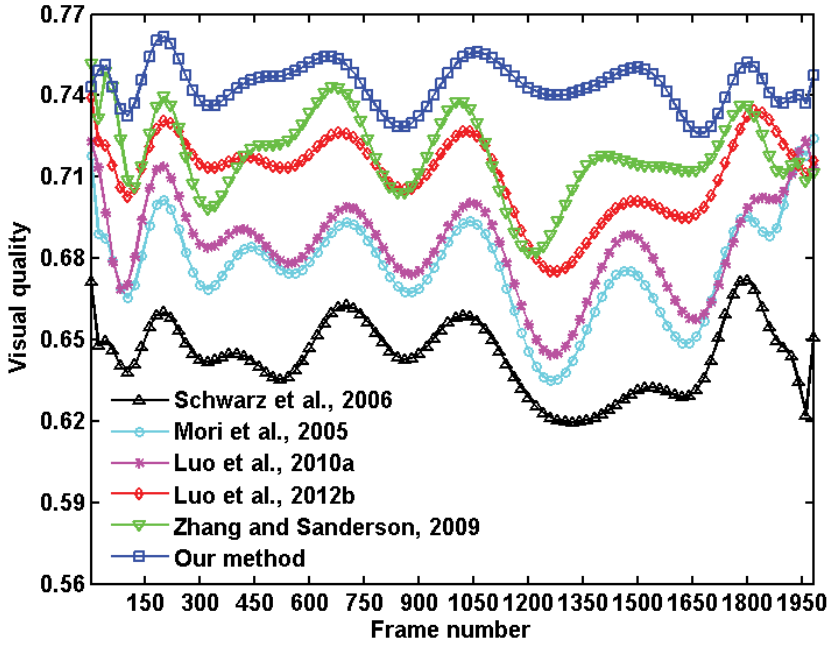


(b) Orientation smoothness

Figure 5.10: Plotted tracking smoothness of six methods evaluated in Experiment 6.

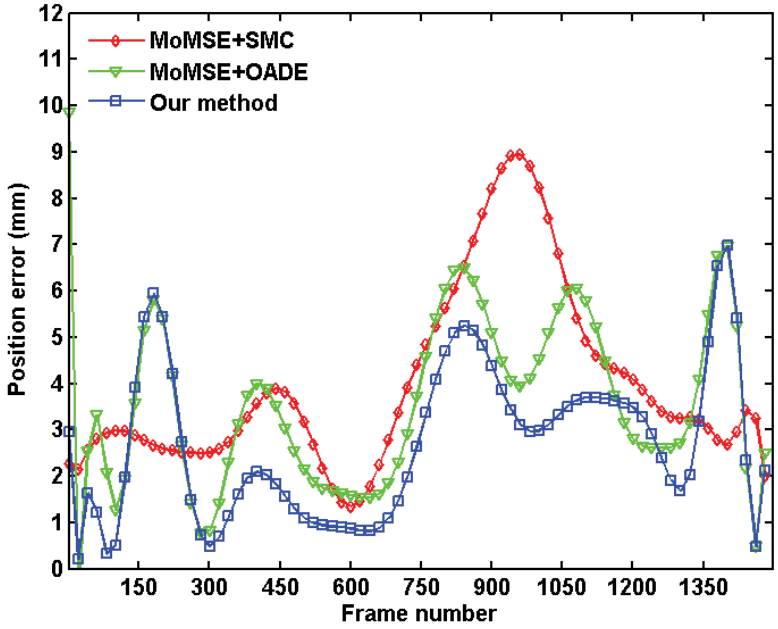


(a)

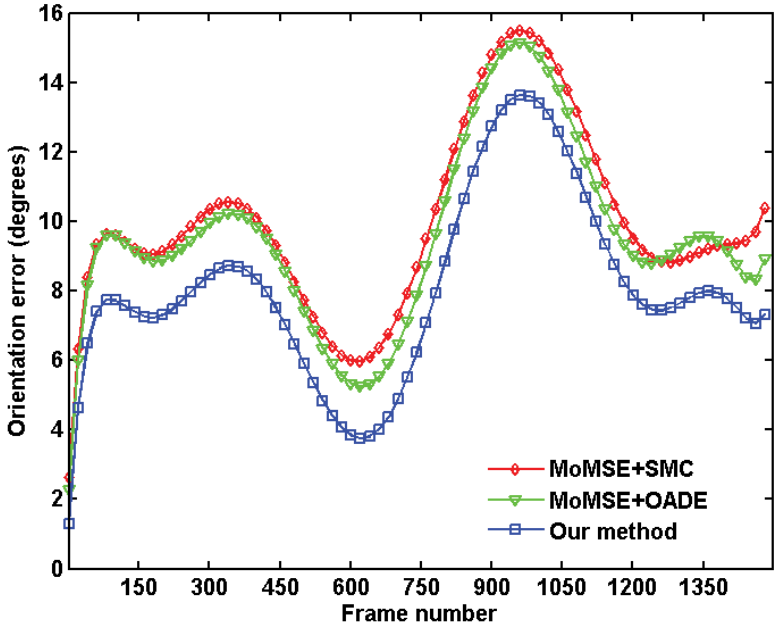


(b)

Figure 5.11: Plotted visual quality of six methods evaluated in Experiments 8 and 14.

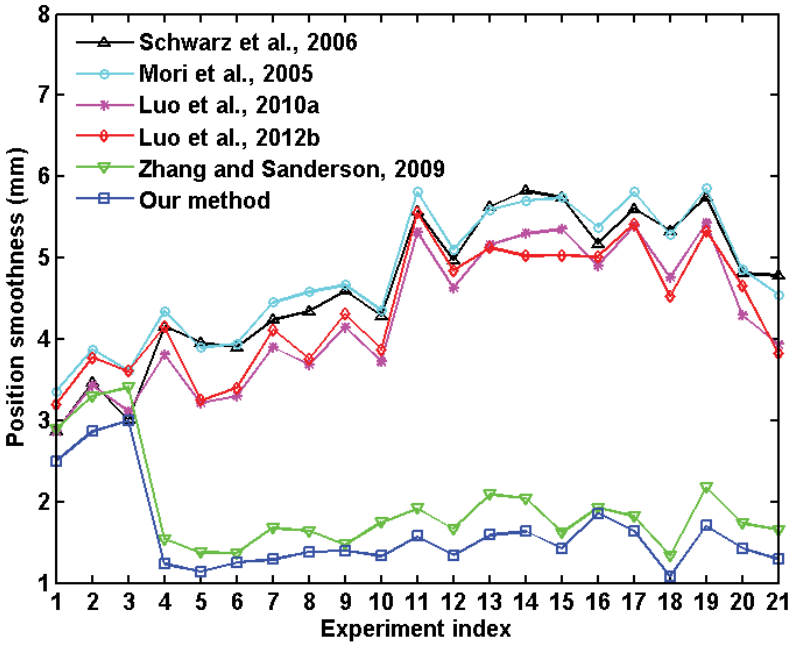


(a) Position error

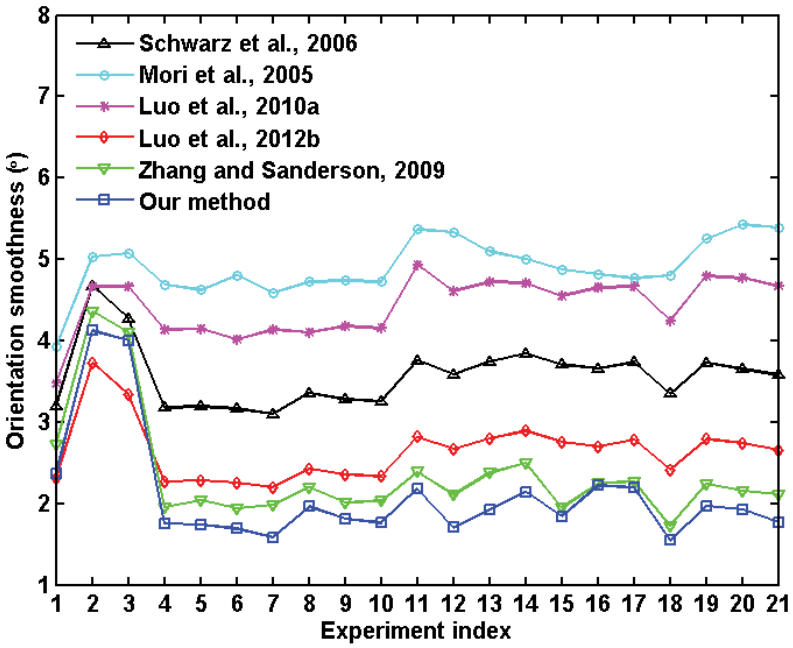


(b) Orientation error

Figure 5.12: Comparison of different similarity measures and optimization frameworks for 3-D bronchoscope motion estimation in Experiment 9. Tracking error of the MoMSE+SMC method [3] was 3.95 mm and 9.99° while the MoMSE+OADE approach was 3.50 mm and 9.58°. Our method shows best performance with 2.69 mm and 8.09°.

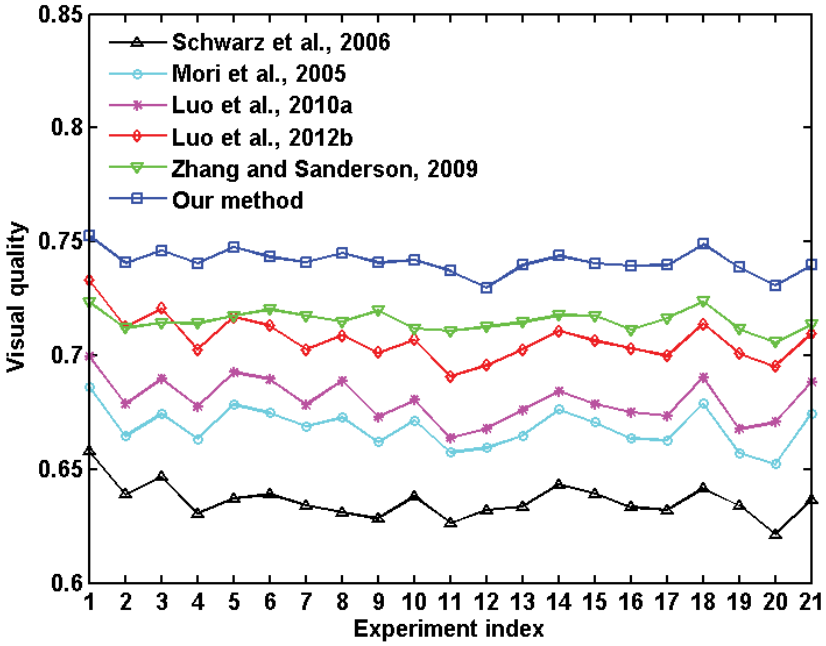


(a) Position smoothness

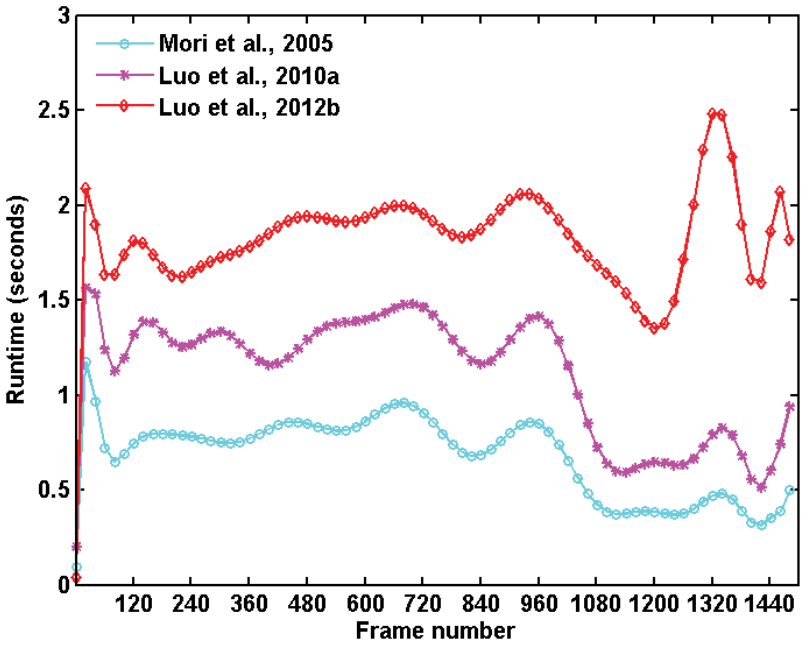


(b) Orientation smoothness

Figure 5.13: Comparison of tracking (position and orientation) smoothnesses of six methods during 21 experiments.

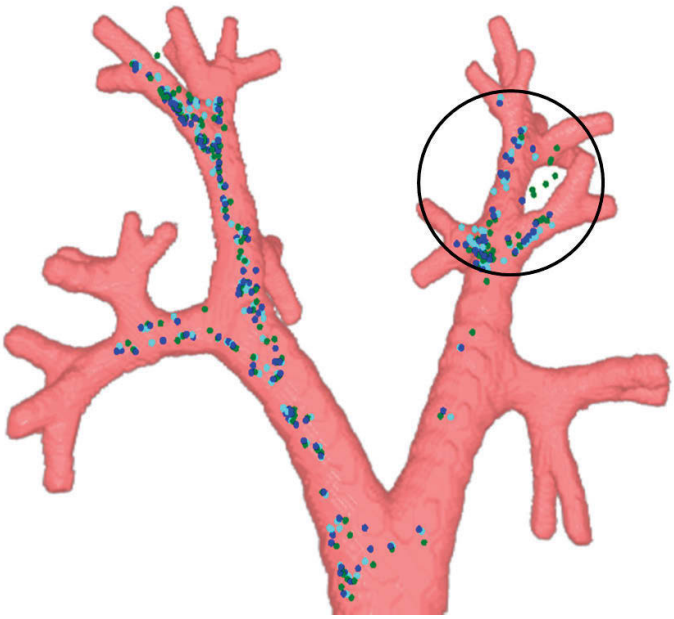


(a) Visual quality

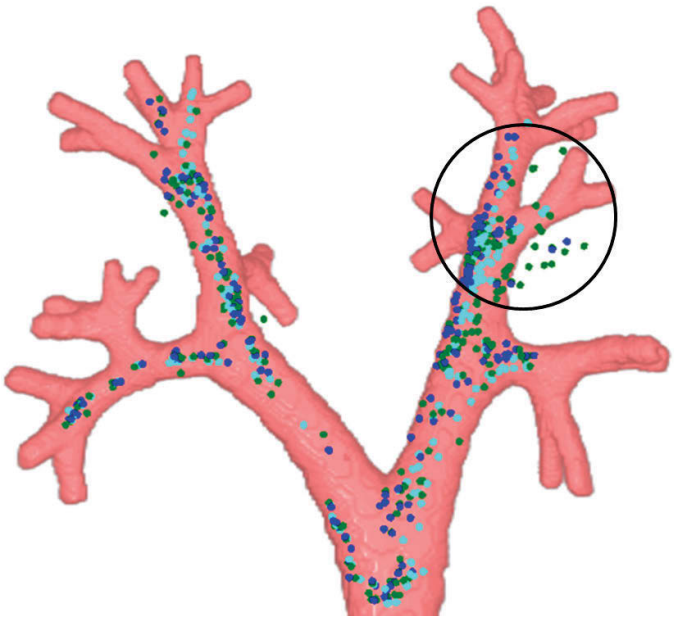


(b) Computational time

Figure 5.14: Visual quality of six methods during 21 experiments and computational time of five methods.



(a) Experiment 5



(b) Experiment 21

Figure 5.15: Plotted tracking results (Experiments 5 & 21) as camera motion paths on pre-built 3-D bronchial tree model using the OADE method (*blue dots*) and the method of Luo et al. [3] (*green dots*). Dots demonstrate that our method overlaps more ground truth dots or follow longer camera movement paths than Luo et al. [3] (*cyan dots* show ground truth). The errors of Luo et al. [3] were about 3.92 mm (Experiment 5) and 4.23 mm (Experiment 21). However, our method provided the tracking errors about 2.78 mm (Experiment 5) and 3.32 mm (Experiment 21). Dots in the *circle* marks further prove that our method outperforms Luo et al. [3], since many *green* dots were located outside the bronchial tree.

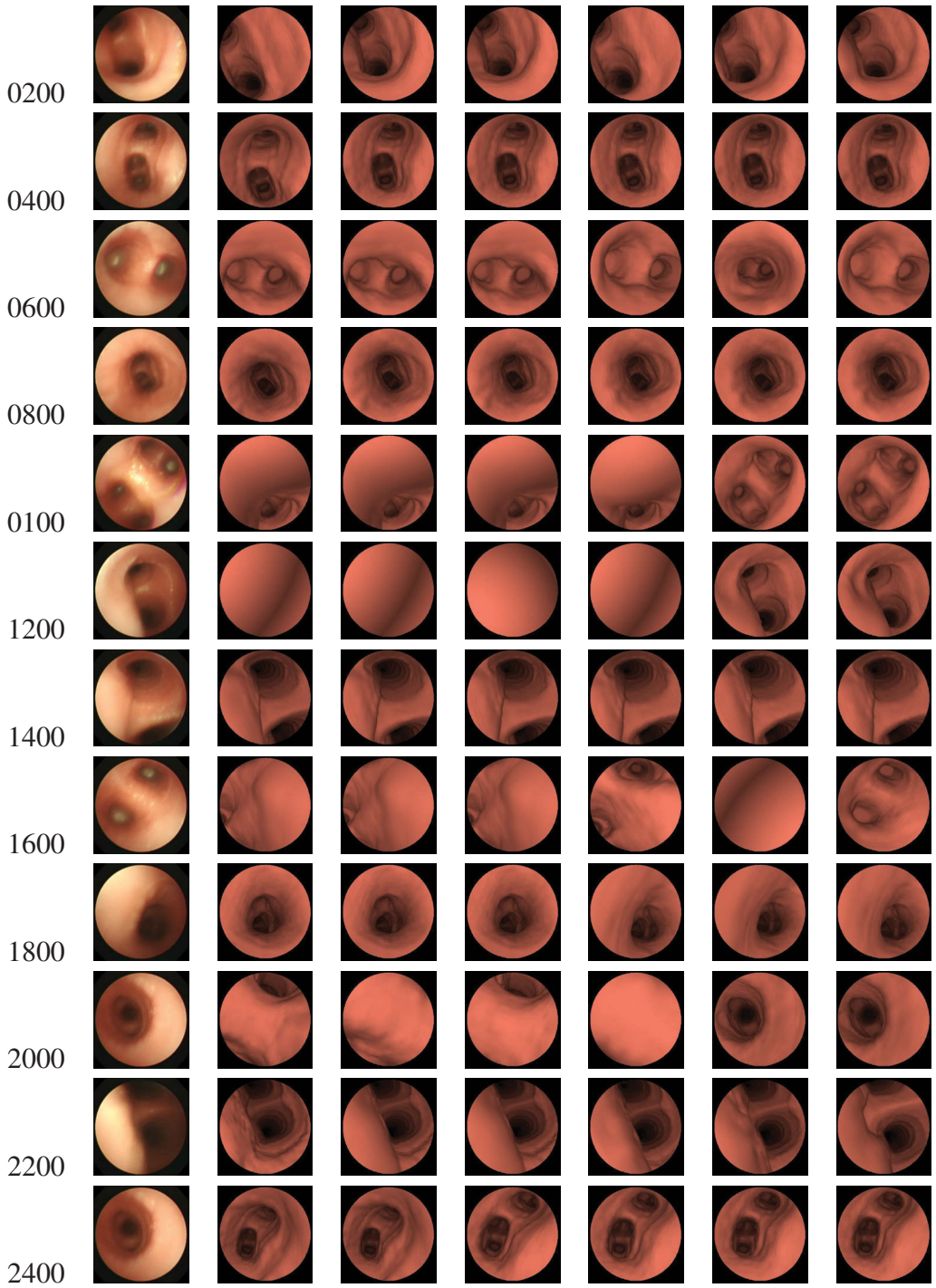


Figure 5.16: Examples of visual comparison of tracking results of Experiment 18. First column shows frame numbers selected uniformly every 200 frames, and second column shows corresponding real images. Other columns (from third to sixth) display virtual images generated from tracking results using methods of [4], [5], [6], [3], [7], and ours, respectively. Our proposed framework shows the best performance.

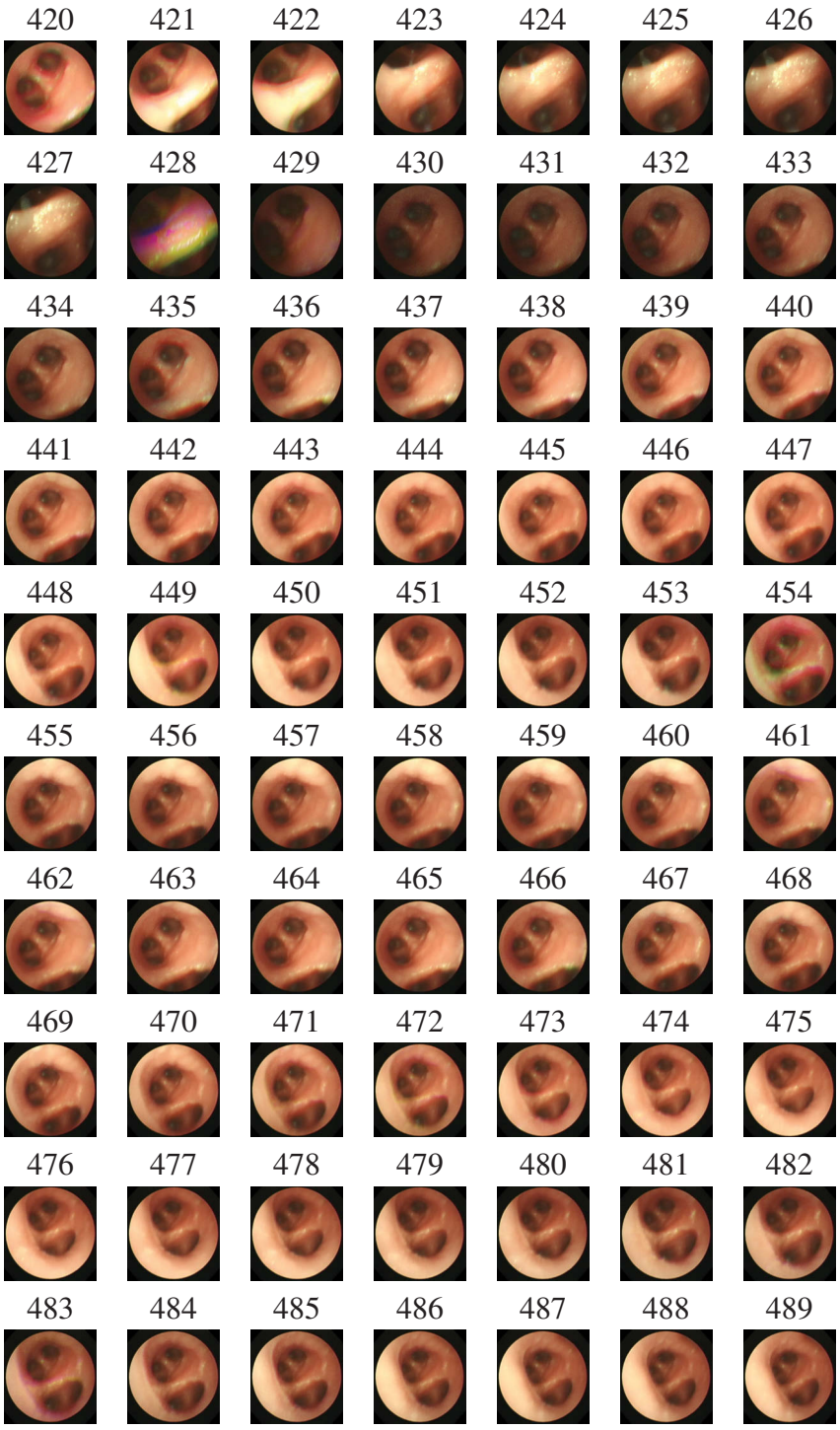


Figure 5.17: Continuous bronchoscopic video frames 420∼489 ($left \longrightarrow right$, $top \longrightarrow bottom$) of Experiment 7 for illustrating hat our method can tackle respiratory motion and image artifacts. Image artifacts, such as illumination changes (e.g., Frames 420, 421, and 422), collisions with bronchial walls (e.g., Frames 423∼427), and jumps (e.g., Frames 428 and 49), occurred frequently.

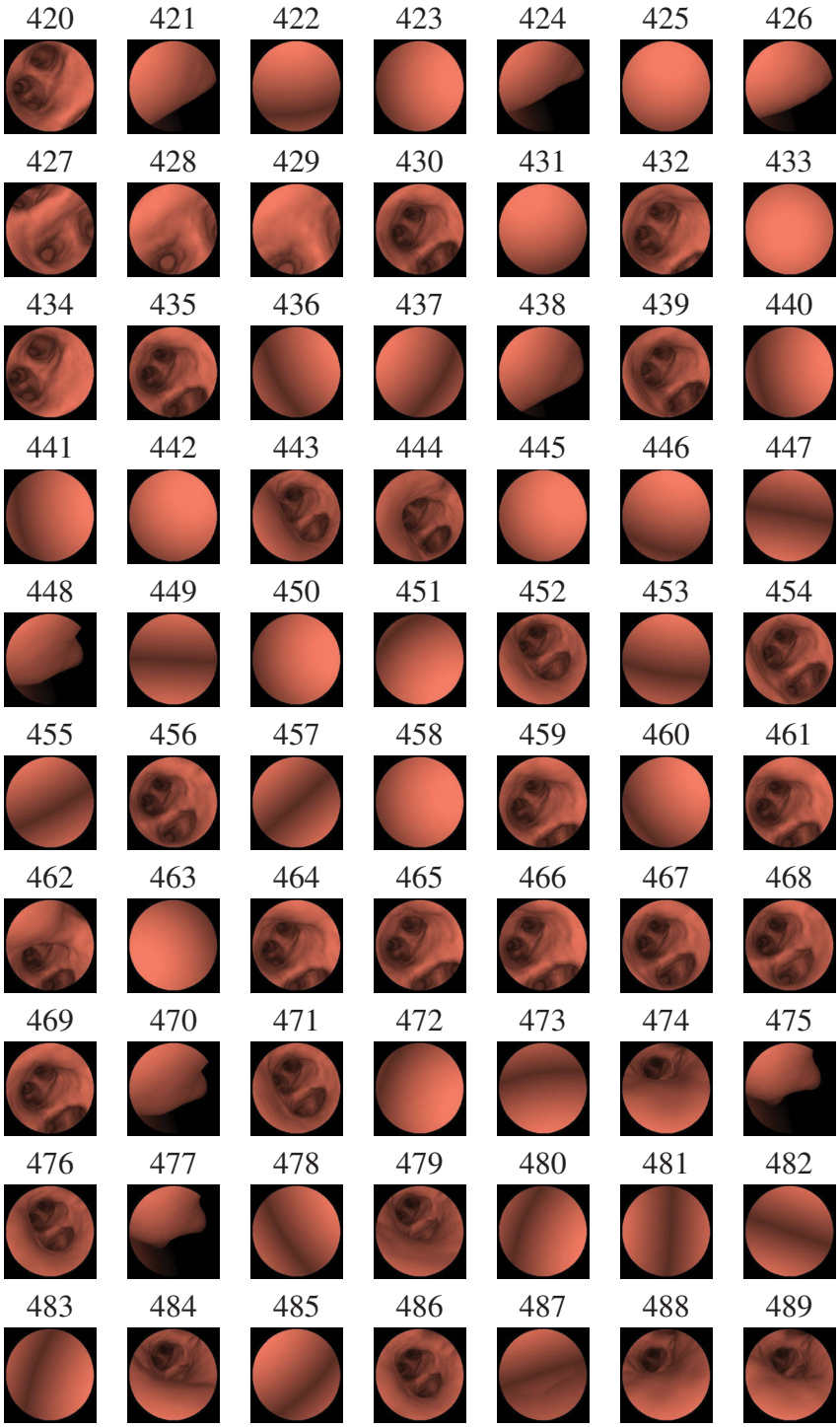


Figure 5.18: Virtual bronchoscopic images that were estimated from the method of Luo et al. [3] correspond to bronchoscopic video images in Fig. 5.17.

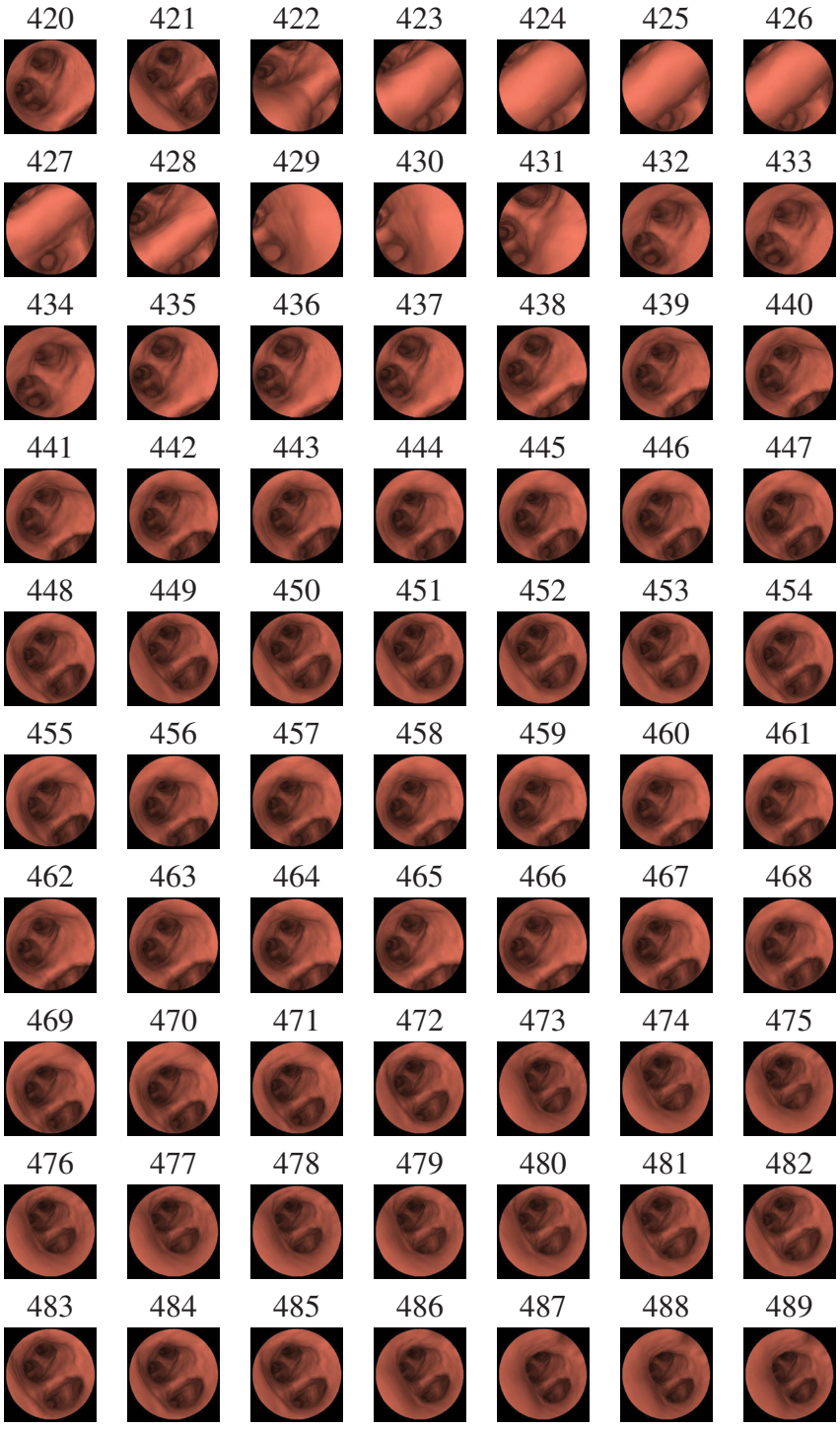


Figure 5.19: Virtual bronchoscopic images that were estimated from our proposed OADE method correspond to bronchoscopic video images in Fig. 5.17. Our method still tracks bronchoscope motions successfully and greatly outperforms the method of Luo et al. [3] since these virtual images resemble video images much better than in Fig. 5.18.

Chapter 6

Conclusions and open remarks

This chapter is to summarize the contributions or accomplishments that have been achieved in this dissertation. The effectiveness of the proposed evolutionarily computed endoscopic navigation frameworks will be discussed, followed by addressing the potential limitations of the proposed approaches. Furthermore, some promising directions and open issues for future work will be also discussed for the development of endoscopic navigation systems.

6.1 Conclusions

This dissertation has been focused on developing a framework of accurate and robust endoscopic navigation systems for clinical applications, particularly image-guided endoscopic interventions. To address some issues that are existed in currently available endoscope tracking and navigation methods, this work proposed an evolutionarily computed endoscopic navigation framework to improve the accuracy and robustness by modifying and improving standard particle swarm optimization and differential evolution algorithms, respectively.

In general, the evolutionary computing-based strategy provides a more accurate and smooth tracking and navigation framework than other available approaches to estimate the endoscopic camera 3D movement and localize the endoscope tip or other channel-passed surgical tools (e.g., biopsy needles) at target regions. The experimental results demonstrate that the proposed frameworks successfully deal with the problems of endoscopic image artifacts (e.g., resulting from specular or inter-reflection, motion blurring, and illumination changes) and inaccurate electromagnetic sensor measurements because of patient movements (e.g., cough), respiratory motion and magnetic field distortion (due to metal materials in the magnetic working volume) of electromagnetic tracking systems during endoscope tracking and navigation. Several technical highlights or achievements are clarified as follows:

1. An adaptive marker-free registration method to spatially align or synchronize the electromagnetic tracker (or sensor) and the pre-clinical image coordinate systems. This approach originally employed a multiple point selection strategy to release the assumption of moving the endoscope along the centerlines of the airway tree during endoscopic intervention. Such a selection strategy has been demonstrated to be very useful to significantly improve the alignment or registration accuracy. **(Chapter 3)**

2. An enhanced particle swarm optimization method was proposed in this thesis. It introduces temporal information (i.e., the current endoscopic video image) to automatically or adaptively compute evolutionary parameters in the step of particle swarm propagation. Compared to pre-defined evolutionary parameters, adaptively determining them can avoid getting trapped in local minima during the optimization procedure. With application to fuse multimodal information including preoperative information (e.g., computed tomography images), endoscopic video sequences, and positional sensor measurements, such an improved optimization provides an accurate and robust endoscope tracking and navigation framework for medical interventions. **(Chapter 4)**

3. An observation-driven adaptive differential evolution algorithm was developed in this work. The performance of the mutation in differential evolution plays an important role in optimization. The current observation in dynamic systems is very important information for any stochastic processing. Based on that, an observation-driven mutation operator was proposed to enhance the performance of differential evolution. Generally speaking, the proposed adaptive differential evolution contains two main advantages that are distinguished from other adaptive differential evolution approaches: (1) the current observation information including the current electromagnetic sensor measurement and the current endoscopic c video image was employed both in the mutation operator and the individual's fitness computation, and (2) the mutation factors and the crossover rate that were used in the individual propagation are also determined adaptively on the basis of the currently observed endoscopic image. **(Chapter 5)**

4. In addition, this dissertation shows the first study on using evolutionary computation for medical application, particularly for image-guided endoscopic interventions. In this respect, firstly extending evolutionary computing algorithms to medical or clinical applications is also one of the main contributions of this PhD research work.

6.2 Benefits to other fields

Although this thesis has been concentrated on medical endoscopic camera3D motion tracking and navigating the surgical tools during computer assisted endoscopic interventions, the enhanced evolutionary computation algorithms proposed in this dissertation still can be extended to apply to 3D camera motion estimation in some other related fields, e.g., robotic system for localizing the robot in its working environments and navigate where to move. Intelligent transportation systems can also benefit from this PhD work as such systems also require to determine the car position and orientation by processing vehicle camera image sequences. On the other hand, the improved optimization approaches could be introduced in hand-held device such as cell-phone systems. In addition, the enhanced evolutionary computation algorithms also can be applied to any other stochastic optimization problems.

6.3 Potential limitations

Even though the evolutionarily computed endoscope tracking and navigation framework proposed in this thesis outperforms the state-of-the-art image-guided endoscopic intervention approaches, it potentially suffers from several limitations, as discussed as follows.

1. The evolutionarily computed endoscopic navigation frameworks do not guarantee to completely resolve the problem of the working volume distortion of the magnetic field generated in electromagnetic tracking systems. Ferrous metals or conductive material unavoidably exist in either dynamic phantom or clinical setups (e.g., the endoscope itself contains metals at its distal tip). Hence, the magnetic filed distortion is still a open issue in EM tracking systems. One of possible solution to correct the magnetic field distortion is to introduce optical tracking techniques in electromagnetic localizer systems, i.e., construct a hybrid magneto-optical tracking and navigation method to assist physicians to realize precise and stable image-guided endoscopic interventions.

2. Respiratory motion is one of big challenges for most image-guided surgical interventions. Removing the influence on the tracking error from breathing motion was not addressed in this work. Modeling respiratory motion to compensate the tracking accuracy and reduce the surgical risk is another open issue in image-guided interventions.

3. One of main purposes of this work is to improve continuous tracking accuracy and meet with the clinical requirement of 2.0 – 4.0 mm during image-guided endoscopic interventions. All the experiments in this work were performed on the basis of the constructed dynamic phantom with simulated respiratory motion. The current tracking

error was validated about 2.89 mm. However, such a tracking error might become larger than evaluating the proposed approaches on clinical or patient datasets. Therefore, further improving the tracking accuracy and patient data validation are the future work in the development electromagnetically navigated endoscopic interventions.

4. On the other hand, structural information extraction from endoscopic video sequences is important for fitness computation, which also challenges the performance of the evolutionarily computed endoscopic navigation approaches. A relatively simple image processing method, which employs hue-saturation-lightness (HSL) model-based color information detection, was implemented to identify the structural regions in accordance with several predefined thresholds. Such a HSL-driven extraction possibly results in inaccurate fitness calculation. Inter-pixel similarity will be introduced to precisely detect the structural regions on endoscopic video images. Moreover, besides the HSL-based extraction approach, more robust functions without defining any thresholds should be considered to perform patching or aggregation in the future work.

5. Another limitation lies in the particle fitness computation. Such a fitness calculation relies heavily on the quality of endoscopic video images. Hence, it could be incorrectly computed and evaluated during particle propagation when image artifacts are observed on endoscopic videos. Incorrect particle fitness computation leads to the weak exploitation capacity of the population in evolutionary computing. To tackle the incorrect fitness computation problem and obtain the powerful exploitation capacity, a more robust image similarity function, which should be greatly insensitive to illumination changes or other image artifacts, will be explored to further improve the tracking accuracy for image-guided endoscopic interventions or examinations in the future.

6. Additionally, patient data validation was not performed in this dissertation. Such clinical evaluation is certainly necessary and undoubtedly optimal. Unfortunately, obtaining clinical data is really difficult since limited clinical data exist worldwide.

6.4 Promising directions

Although evolutionary computed endoscope tracking and navigation provides some significantly improved points, compared to some currently available approaches, some challenges still remain in this field. To make evolutionary computed endoscopic navigation more powerful, several promising research directions are shown in the following.

1. **Uncertainty detection.** Uncertain measurements are of major obstacles for continuously tracking the endoscope movements. Surgical uncertainty refers mainly to problematic or uninformative video frames, resulting from severe endoscope shake, sudden patient cough, or image prolonged occlusions (e.g., caused by pulmonary mucus of bubbles or blood), which are commonly occurred during endoscopic interventions. Such uncertainty absolutely leads to erratic camera motion estimation from the point of view of the computer vision field. In particular, the endoscopic video image uncertainty is completely unexpected and unpredictable particularly when they appear in a long sequence (many consecutive frames). The development of uncertainty detection and removal is an promising direction in endoscopic navigation. is likely to be weak and insufficient to diminish uncertainty. To further improve the proposed approach's robustness to uncertain measurements or erratic motions, it is indispensable to explore a mechanism of uncertainty detection and removal in image-guided endoscopic interventions, e.g., using a manifold learning method to identify uncertain measurements before the video-volume registration for navigating the endoscope [14].

2. **Respiratory motion compensation.** Patient movements, particularly respiratory motion, usually deteriorates the tracking accuracy, since they might cause inaccurate sensor measurements that may not exactly correspond to the current endoscope position. Compensating respiratory motion for accurate endoscope tracking and navigation is another interesting research Based on 4D CT data that are consistently recorded alongside the patient's natural breathing motion, constructing dynamic virtual endoscopy is important to improve the accuracy and robustness of surgical tool movement tracking, localization, and navigation for medical procedure applications in operating rooms [14].

3. **Ultrasound-guided endoscopic navigation.** Endobronchial ultrasound (EBUS) is a general innovative endoscopic diagnostic technique [12, 67]. It enables direct visualization of the multilayer structures of the bronchial wall and the adjacent mediastinal structures. However, EBUS alone lacks the ability to navigate the endoscope to target regions and requires the ednsocopist to operate the endoscope blindly to peripheral lung lesions. Therefore, the development of an EBUS navigation system to guide physicians in endoscopy is also necessary research direction in the field of computer assisted endoscopic intervention. On the other hand, the diameter of the EBUS tip is quite big since EBUS usually are integrated with a convex ultrasound probe. Such a big-size tip limits the EBUS's, particularly it cannot go through the peripheral bronchial

branches. Using and navigating miniature radial probes through the working channel of the endoscope to augment endoscopic interventions is also a promising direction.

4. **Simultaneous localization and mapping (SLAM).** The SLAM techniques are widely used in robotic system to localize robot itself at a scenario where the robot walks or perform any tasks. They are basic algorithms for camera tracking and navigation. The sensor-based tracking approaches suffer from the jitter, drift or expensive computation because of the characteristics of individual sensor system. Fortunately, extending SLAM to sensor-based tracking or navigating algorithms is effective way to deal with these issues. In electromagnetic sensor-based endoscope tracking and and navigation, the electromagnetic sensor's measurements unavoidably involve static and jitter errors. The jitter or dynamic error, which remains influenced by the magnetic field distortion of the electromagnetic tracking systems, is difficult to be corrected. Introducing SLAM techniques into electromagnetically navigated endoscopic interventions is a promising way to solve the problem of inaccurate electromagnetic sensor measurement with jitter or dynamic error, potentially resulting in the improvement of the tracking accuracy.

5. **Microscopic augmentation.** Online pathological analysis is very interesting by surgeons in operating rooms. Unfortunately, either current conventional or image-guided endoscopic procedures has no such a pathological analysis function. Combining endoscopy with microscopy is a potential solution in on-site endoscopic interventions.

6.5 Closing remarks

This dissertation aimed at developing a framework of accurate and robust endoscopic navigation for clinical applications. The concept of evolutionary computation was introduced into image-guided endoscopic interventions. Two evolutionary computing algorithms of particle swarm optimization and differential evolution were modified to improve their performance when using them in any stochastic optimization problems. With their applications to endoscope tracking and navigation, an evolutionarily computed endoscopic navigation framework was proposed and demonstrated to provide more accurate and robust image-guided endoscopic interventions, compared to currently available approaches. Unfortunately, there are still some challenges necessary to be addressed on endoscopic navigation in the future.

Eventually, this dissertation would conclude that computer assisted endoscopic navigation and interventions are developing under a more interdisciplinary domain involved in those academic and professional disciplines with collaboration in the following fields: *(1) Image Engineering*, *(2) Pattern Recognition*, *(3) Computer Graphics*, *(4) Sensor Technology*, *(5)*

Computer Vision, and *(6) Information Fusion Theory*. The integration of the advantages of these techniques would certainly revolutionize image-guided endoscopic interventions.

References

[1] Daisuke Deguchi, Marco Feuerstein, Takayuki Kitasaka, Yasuhito Suenaga, Ichiro Ide, Hiroshi Murase, Kazuyoshi Imaizumi, Yoshinori Hasegawa, and Kensaku Mori. Real-time marker-free patient registration for electromagnetic navigated bronchoscopy: a phantom study. *International Journal of Computer Assisted Radiology and Surgery*, 7(3):359–369, 2012.

[2] Tassilo Klein, Joerg Traub, Hubert Hautmann, Alireza Ahmadian, and Nassir Navab. Fiducial-free registration procedure for navigated bronchoscopy. In *Proc. MICCAI'07*, volume LNCS 4791, pages 475–482, 2007.

[3] X. Luo, M. Feuerstein, T. Kitasaka, and K. Mori. 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(3):371–387, 2012.

[4] Y. Schwarz, J. Greif, H. D. Becker, A. Ernst, and A. Mehta. Real-time electromagnetic navigation bronchoscopy to peripheral lung lesions using overlaid CT images: the first human study. *Chest*, 129(4):988–994, 2006.

[5] K. Mori, D. Deguchi, K. Akiyama, T. Kitasaka, C. R. Maurer Jr., Y. Suenaga, H. Takabatake, M. Mori, and H. Natori. Hybrid bronchoscope tracking using a magnetic tracking sensor and image registration. In *Proc. MICCAI'05*, volume 3750, pages 543–550, 2005.

[6] Xiongbiao Luo, Marco Feuerstein, Takayuki Kitasaka, Kazuyoshi Imaizumi, Yoshinori Hasegawa, and Kensaku Mori. Towards hybrid bronchoscope tracking under respiratory motion: evaluation on a dynamic motion phantom. In *Proc. SPIE MI 2010*, volume 7625, page 76251B, 2010.

[7] J. Zhang and Arthur C. Sanderson. JADE: Adaptive differential evolution with optional external archive. *IEEE Transactions on Evolutionary Computation*, 13(5):945–958, 2009.

[8] X. Luo, T. Reichl, M. Feuerstein, T. Kitasaka, and K. Mori. Modified hybrid bronchoscope tracking based on sequential monte carlo sampler: Dynamic phantom validation. In *Proc. ACCV 2010*, volume LNCS 6494, pages 409–421, 2010.

[9] X. Zhang, W. Hu, W. Qu, and S. Maybank. Multiple object tracking via species-based particle swarm optimization. *IEEE Transactions on Circuits and Systems for Video Technology*, 22:1590–1602, 2010.

[10] American Cancer Society. Global cancer facts & figures 3rd edition. Atlanta: American Cancer Society, 2015.

[11] Pallav L Shah. Flexible bronchoscopy. *Medicine*, 36(3):151–154, 2008.

[12] Armin Ernst. *Introduction to bronchoscopy*. Cambridge University Press, 2009.

[13] Yang Xia and Ko-Pen Wang. Transbronchial needle aspiration: where are we now? *Journal of thoracic disease*, 5(5):678, 2013.

[14] Xiongbiao Luo. *New approaches to endoscopic camera motion tracking for bronchoscopy navigation*. PhD thesis, Nagoya University, 2011.

[15] Felix JF Herth, Ralf Eberhardt, and Armin Ernst. The future of bronchoscopy in diagnosing, staging and treatment of lung cancer. *Respiration*, 73(4):399–409, 2006.

[16] A. J. Chung, F. Deligianni, P. Shah, A. Wells, and G. Z. Yang. Patient-specific bronchoscopy visualization through BRDF estimation and disocclusion correction. *IEEE Transactions on Medical Imaging*, 25(4):503–513, 2006.

[17] James P Helferty, AJ Sherbondy, Atilla P Kiraly, and William E Higgins. Computer-based system for the virtual-endoscopic guidance of bronchoscopy. *Computer Vision and Image Understanding*, 108(1):171–187, 2007.

[18] X. Luo, M. Feuerstein, D. Deguchi, T. Kitasaka, H. Takabatake, and K. Mori. Development and comparison of new hybrid motion tracking for bronchoscopic navigation. *Medical Image Analysis*, 16(3):577–596, 2012.

[19] L. Y. Ching, K. Moller, and J. Suthakorn. Non-radiological colonoscope tracking image guided colonoscopy using commercially available electromagnetic tracking system. In *Proc. IEEE Conf. on Robotics Automation and Mechatronics'10*, pages 62–67, 2010.

[20] A. Joshi, D. Scheinost, R. Globinsky, K. P. Vives, D. D. Spencer, L. H. Staib, and X. Papademetris. Augmented inline-based navigation for stereotactic image guided neurosurgery. In *Proc. IEEE ISBI'11*, pages 1869–1872, 2011.

[21] D.J. Mirota, H. Wang, R.H. Taylor, M. Ishii, G.L. Gallia, and G.D. Hager. A system for video-based navigation for endoscopic endonasal skull base surgery. *IEEE Transactions on Medical Imaging*, 31(4):963–76, 2012.

[22] Marc Levoy. A hybrid ray tracer for rendering polygon and volume data. *IEEE Computer Graphics and Applications*, 10(2):33–40, 1990.

[23] U. Tiede, K. H. Hoehne, M. Bomans, A. Pommert, M. Riemer, and G. Wiebecke. Investigation of medical 3D-rendering algorithms. *IEEE Computer Graphics and Applications*, 10(2):41–53, 1990.

[24] D. Deguchi, K. Mori, M. Feuerstein, T. Kitasaka, C. R. Maurer Jr., Y. Suenaga, H. Takabatake, M. Mori, and H. Natori. Selective image similarity measure for bronchoscope tracking based on image registration. *Medical Image Analysis*, 13(4):621–633, 2009.

[25] T. D. Soper, D. R. Haynor, R. W. Glenny, and E. J. Seibel. In vivo validation of a hybrid tracking system for navigation of an ultrathin bronchoscope within peripheral airways. *IEEE Transactions on Biomedical Engineering*, 57(3):736–745, 2010.

[26] M. Isard and A. Blake. Condensation - conditional density propagation for visual tracking. *International Journal of Computer Vision*, 29(1):5–28, 1998.

[27] I. Gergel, T. R. dos Santos, R. Tetzlaff, L. Maier-Hein, H.-P. Meinzer, and I. Wegner. Particle filtering for respiratory motion compensation during navigated bronchoscopy. In *Proc. SPIE MI'10*, volume 7625, page 76250W, 2010.

[28] F. Deligianni, A. J. Chung, and G. Z. Yang. Nonrigid 2-D/3-D registration for patient specific bronchoscopy simulation with statistical shape modeling: Phantom validation. *IEEE Transactions on Medical Imaging*, 25(11):1462–1471, 2006.

[29] M. Feuerstein, T. Kitasaka, and K. Mori. Adaptive branch tracing and image sharpening for airway tree extraction in 3-d chest ct. In *2nd Int. Workshop on Pulmonary Image Analysis*, pages 273–284, 2009.

[30] H. D. Becker, F. Herth, A. Ernst, and Y. Schwarz. Bronchoscopic biopsy of peripheral lung lesions under electromagnetic guidance: a pilot study. *Journal of Bronchology and Interventional Pulmonology*, 12(1):9–13, 2005.

[31] D. Makris, A. Scherpereel, S. Leroy, B. Bouchindhomme, J. B. Faivre, J. Remy, P. Ramon, and C. H. Marquette. Electromagnetic navigation diagnostic bronchoscopy for small peripheral lung lesions. *European Respiratory Journal*, 29(6):1187–1192, 2007.

[32] L. M. Seijo, J. P. de Torres, M. D. Lozano, G. Bastarrika, A. B. Alcaide, M. M. Lacunza, and J. J. Zulueta. Diagnostic yield of electromagnetic navigation bronchoscopy is highly dependent on the presence of a bronchus sign on CT imaging: results from a prospective study. *Chest*, 138(6):1316–1321, 2010.

[33] I. Khan, R. Chin, N. Adair, A. Chatterjee, E. Haponik, and J. Conforti. Electromagnetic navigation bronchoscopy in the fiagnosis of peripheral lung lesions. *Clinical Pulmonary Medicine*, 18(1):42–45, 2011.

[34] A.E. Eiben and J.E. Smith. *Introduction to Evolutionary Computing*. Springer-Verlag Berlin Heidelberg, 2003.

[35] Kenneth A De Jong. *Evolutionary Computation : A Unified Approach*. MIT Press, 2006.

[36] R. Chiong, Th. Weise, and Z. Michalewicz. *Variants of Evolutionary Algorithms for Real-World Applications, Springer, 2012, ISBN 3642234232*. Springer-Verlag Berlin Heidelberg, 2012.

[37] Richard Hartley and Andrew Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2004.

[38] I. Bricault, G. Ferretti, and P. Cinquin. Registration of real and CT-derived virtual bronchoscopic images to assist transbronchial biopsy. *IEEE Transactions on Medical Imaging*, 17(5):703–714, 1998.

[39] Z. Y. Zhang. A flexible new technique for camera calibration. *IEEE Transations on Pattern Analysis and Machine Intelligence*, 22(11):1330–1334, 2000.

[40] R. Y. Tsai and R. K. Lenz. A new technique for fully autonomous and efficient 3d robotics hand/eye calibration. *IEEE Transactions on Robotics and Automation*, 5(3):345–358, 1989.

[41] D.W. Eggert, A. Lorusso, and R.B. Fisher. Estimating 3-D rigid body transformations: a comparison of four major algorithms. *Machine Vision and Applications*, 9(5-6):272–290, 1997.

[42] Xiongbiao Luo and Kensaku Mori. Adaptive fiducial-free registration usingmultiple point selection for real-time electromagneticallynavigated endoscopy. In *Proceedings of the SPIE MI'14: Image-Guided Procedures, Robotic Interventions, and Modeling*, volume 9036, 2014.

[43] F. V. Berghen and H. Bersini. CONDOR, a new parallel, constrained extension of powell's UOBYQA algorithm: experimental results and comparison with the DFO algorithm. *Journal of Computational and Applied Mathematics*, 181(1):157–175, 2005.

[44] J. Kennedy and R. C. Eberhart. Particle swarm optimization. In *Proc. IEEE Int. Conf. on Neural Networks'95*, pages 1942–1948, 1995.

[45] S. Mostaghim and J. Teich. Strategies for finding good local guides in multi-objective particle swarm optimization. In *Proc. IEEE ISI'03*, pages 26–33, 2003.

[46] B. Xue, M. Zhang, and W. N. Browne. Particle swarm optimization for feature selection in classification: A multi-objective approach. *IEEE Transactions on Cybernetics*, 43(6):1656–1671, 2013.

[47] D. Parrott and X. Li. Locating and tracking multiple dynamic optima by a particle swarm model using speciation. *IEEE Transactions on Evolutionary Computation*, 10(4):440–458, 2006.

[48] L. Zhang, T. Mei, Y. Liu, D.Tao, and H. Q. Zhou. Visual search reranking via adaptive particle swarm optimization. *Pattern Recognition*, 44(8):1811–1820, 2011.

[49] David Salomon. *The Computer Graphics Manual*. Springer, 2011.

[50] M. S. Arulampalam, S. Maskell, and N. Gordon. A tutorial on particle filters for nonlinear/non-gaussian Bayesian tracking. *IEEE Transaction on Signal Processing*, 50:174–188, 2002.

[51] R. Storn and K. V. Price. Differential evolution - A simple and efficient heuristic for global optimization over continuous spaces. *Journal of Global Optimization*, 11(4):341–359, 1997.

[52] S. Das and P. N. Suganthan. Differential evolution: A survey of the state-of-the-art. *IEEE Transactions on Evolutionary Computation*, 15(1):4–31, 2011.

[53] A. K. Qin, V. L. Huang, and P. N. Suganthan. Differential evolution algorithm with strategy adaptation for global numerical optimization. *IEEE Transactions on Evolutionary Computation*, 13(2):398–417, 2009.

[54] Ernesto Mininno, Ferrante Neri, Francesco Cupertino, and David Naso. Compact differential evolution. *IEEE Transactionson Evolutionary Computation*, 15(1):32–54, 2011.

[55] Massimiliano Vasile, Edmondo Minisci, and Marco Locatelli. An inflationary differential evolution algorithm for space trajectory optimization. *IEEE Transactionson Evolutionary Computation*, 15(2):267–281, 2011.

[56] Erik B. Dam, Martin Koch, and Martin Lillholm. Quaternions, interpolation and animation. Technical Report DIKU-TR-98/5, University of Copenhagen, Denmark, 1998.

[57] J. Liu and J. Lampinen. A fuzzy adaptive differential evolution algorithm. *Soft Computing - A Fusion of Foundations, Methodologies and Applications*, 9(6):448–462, 2005.

[58] J. Brest, S. Greiner, B. Boskovic, M. Mernik, and V. Zumer. Self-adapting control parameters in differential evolution: A comparative study on numerical benchmark problems. *IEEE Transactions on Evolutionary Computation*, 10(6):646–657, 2006.

[59] J. Teo. Exploring dynamic self-adaptive populations in differential evolution. *Soft Computing - A Fusion of Foundations, Methodologies and Applications*, 10(8):673–686, 2006.

[60] A. Doucet, S. Godsill, and C. Andrieu. On sequential Monte Carlo sampling methods for Bayesian filtering. *Statistics and Computing*, 10(3):197–208, 2000.

[61] K. V. Price, R. M. Storn, and J. A. Lampinen. *Differential evolution: A practical approach to global optimization*. Springer-Verlag, New York, 2005.

[62] E. Mezura-Montes, J. Velazquez-Reyes, and C. A. Coello Coello. Modified differential evolution for constrained optimization. In *Proc. IEEE CEC 2006*, pages 25–32, 2006.

[63] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004.

[64] Xiongbiao Luo and Kensaku Mori. A discriminative structural similarity measure and its application to video-volume registration for endoscope three-dimensional motion tracking. *IEEE Transactions on Medical Imaging*, 33(6):1248–1261, 2014.

[65] Z. Wang and A. C. Bovik. A universal image quality index. *IEEE Signal Processing Letters*, 9(3):81–84, 2002.

[66] Xiongbiao Luo and Kensaku Mori. Observation-driven adaptive differential evolution for robust bronchoscope 3-d motion tracking. In *Proc. ACCV 2012*, volume LNCS 7726, pages 259–271, 2012.

[67] A. R. Haas, A. Vachani, and D. H. Sterman. Advances in diagnostic bronchoscopy. *American Journal of Respiratory and Critical Care Medicine*, 182(5):589–597, 2010.

Appendix A

List of publications

Peer-reviewed journal and conference papers since 2012:

- **Ying Wan***, Qiang Wu, and Xiangjian He. Dense feature correspondence for video-based endoscope three-dimensional motion tracking. *Biomedical and Health Informatics (BHI), 2014 IEEE-EMBS International Conference on*, pp.49–52, Spain.
- Xiongbiao Luo*, **Ying Wan***, Xiangjian He, Jie Yang, and Kensaku Mori. Diversity-enhanced Condensation algorithm and its application for robust and accurate endoscope three-dimensional motion tracking. *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*, pp.1250–1257, USA.
- Xiongbiao Luo*, **Ying Wan***, and Xiangjian He. Robust electromagnetically guided endoscopic procedure using enhanced particle swarm optimization for multimodal information fusion. *Medical Physics* 42(4): 1808–1817, 2015. **ERA Tier A Journal**
- Xiongbiao Luo*, **Ying Wan***, Xiangjian He, and Kensaku Mori. Adaptive marker-free registration using a multiplepoint strategy for real-time and robust endoscope electromagnetic navigation. *Computer Methods and Programs in Biomedicine* 118(2): 147-57, 2015. **ERA Tier A* Journal**
- Xiongbiao Luo*, **Ying Wan***, Xiangjian He, and Kensaku Mori. Observation-driven adaptive differential evolution and its application to accurate and smooth bronchoscope three-dimensional motion tracking. *Medical Image Analysis*, DOI: j.media.2015.01.002. **ERA Tier A* Journal**