Study on bronchoscope tracking for bronchoscopy navigation

Cheng Wang

Abstract

This thesis describes the author's work on the study of image-based bronchoscope tracking for bronchoscopy navigation. In the United States and worldwide, the populations death from lung cancer and bronchial cancer are the most among people who die from other common cancers. Early diagnosis can effectively improve the survival rate of patients because the respiratory system is an essential organ of the human body. There are many diagnosis methods, such as the use of computed tomography (CT) images to diagnose the lesions; or the use of a flexible endoscope for pathological examination of the bronchial lumen (transBronchial Biopsy). However, during transBronchial biopsy, since the field of view (FoV) of the camera is narrow and the obtained real bronchoscopic (RB) images have high similarity in appearance. Bronchoscopist is easy to get lost in the tree-like bronchus. Therefore, a navigation system is used to locate the bronchoscope camera in three-dimensional (3D) bronchus and provide the necessary 3D information to the bronchoscopist during bronchoscopy.

A bronchoscopy navigation system works like a 'car navigation system'. The bronchoscopy navigation system shows the camera position of the bronchoscope in real-time in CT image coordinate to navigate physicians during examination. The kernel part of the navigation is bronchoscope tracking, which is used to estimate the camera pose of the bronchoscope. Existing navigation systems either use video-CT-based tracking or an additional electromagnetic (EM) sensor-based bronchoscope tracking. However, for the

type of video-CT-based tracking, the tracking result suffers from the difference between preoperative CT images and intraoperative images; and the EM sensor-based tracking is easily affected by metallic surgical tools and patient movements. To accurately estimate the camera pose, we use methods in computer vision to precisely estimate the camera pose for each coming RB frame. We use a technique called simultaneous localization and mapping (SLAM) to recover the camera pose of a bronchoscope. SLAM uses image sequence as input. The 2D key points detected in the RB images are used to estimate the camera pose and reconstruct the 3D surroundings. This technique calculates the camera pose by minimizing the reprojection error of the 3D points on the image plane. The original method is mainly designed for ordinary indoor/outdoor scenes, however, the bronchus scene contains more complex scenes such as organ deformation caused by patient breathing. Therefore, we improve the original SLAM before applying it for bronchoscope tracking. The 3D position of the point is considered as a condition to find 3D points, which are used to find accurate points for tracking. For validation, two phantoms are used to create bronchoscopic videos. The deformation from the breath is simulated by adding periodic force to the phantom. Experimental results showed that the proposed method is more accurate and more stable than the original method. The proposed method tracks more continuous frames as well. This method is described in detail in Chapter 2.

The complex bronchus scene leads to additional difficulties in image-guided navigation tasks. Therefore, understanding of the operation field is critical during examination or surgery. To this end, a segmentation method is proposed to extract the bronchial orifice region, which is an important characteristic of the bronchus scene on RB video frames, to assist the physicians. Previous works use image appearance and gradation of the RB image for the segmentation task, which behaves poorly in scenes of existing bubbles or changes in illumination. We use the distance between camera and bronchial anatomical structure, which is represented as depth image, to segment the bronchial orifice region instead of a color image to overcome the previous works' shortcomings. According to the previous literature, depth image-based image-guided procedures perform better than color image-based procedures. Depth image is estimated using an image-to-image translation network named cycle generative adversarial network (CycleGAN). CycleGAN is trained to find the relationship between the RB image domain and the depth image domain. We segment the bronchial orifice region individually for each image. We use the vertical and horizontal projection profile curves from the depth image to decide the threshold value used for the binarization. The nonzero region in the binarized image is considered as the bronchial orifice region. This algorithm is described in detail in Chapter 3.

The existing navigation systems estimate the camera pose precisely for each frame, However, the conventional bronchoscope tracking methods will gradually accumulate tracking errors, which will lead the tracking procedure to fail easily. Therefore, a coarse camera localization method is proposed to roughly localize the camera in the bronchi branch using the changes in the anatomical structure of the bronchi. The bronchial orifice is an important anatomical structure in bronchus and it changes with the level of the bronchus. Therefore, the understanding of the changes in the bronchial orifice will benefit a lot for the navigation system. The changes in the bronchial orifice are used to estimate the current bronchial level. The orifice region is obtained by using depth images, which are generated by an image-to-image translation network named cycle generative adversarial network (CycleGAN). The camera movement direction is obtained from the feature point-based camera pose estimation. The branching level is estimated by considering the changes of the bronchial orifice and the camera moving direction. Experimental results showed that the proposed method could estimate the branching level with high accuracy. This method is described in detail in Chapter 4.

At last, the advantages and disadvantages of precise and coarse bronchoscope tracking are compared. From the comparisons, a conclusion that the characteristics based tracking will benefit the bronchoscope tracking is obtained. This is described in Chapter 5.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

This thesis summarizes the author's work in real bronchoscopic (RB) image-based bronchoscope tracking. In Section 1.1, the current status of lung/bronchus cancer and other respiratory diseases worldwide is outlined; in Section 1.2, several existing methods for the diagnosis of lung/bronchus disease; in Section 1.3, the past and current status of bronchoscopy together with the advantages and disadvantages of bronchoscopy are explained; in Section 1.4, the bronchoscopy navigation system is explained from three aspects: definition, component, and existing navigation systems; in Section 1.5, the core part of the navigation: the bronchoscope tracking, is introduced; in Section 1.6, the positioning and the purpose of this research are explained in detail; and finally, in Section 1.7, the structure of this thesis is illustrated.

1.1 Lung and bronchus cancer

According to recent surveys by the American Cancer Society [1, 2, 5, 6], lung and bronchus cancer occupy the leading causes of death from common cancer. There are about 228,820 estimated new cases in 2020, accounting for approximately 12.7% of all

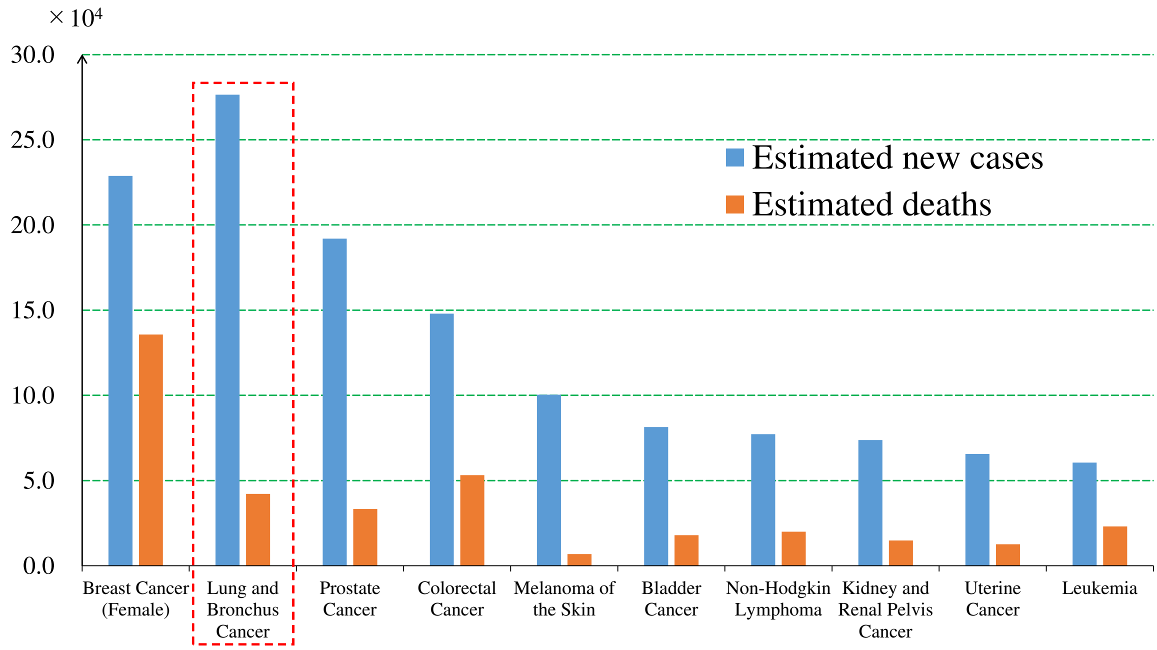


Figure 1.1: Estimated new cases and deaths from common cancers in 2020[1, 2]: estimated deaths from lung and bronchus cancer are in first place.

newly confirmed cases; there were 135,720 estimated deaths in 2020, accounting for approximately 22.4% of deaths from cancer (as shown in Fig. 1.1). Moreover, patients who died from lung/bronchus cancer occupy first place among deaths from cancer. Lung cancer and bronchus cancer are important causes of cancer deaths in the United States and worldwide [7, 8]. The respiratory system is an important organ of the human body, therefore, diseases in this region are more likely to lead to patient death.

There are many causes of lung/bronchus diseases, among which smoking and secondhand smoke are the most critical. Moreover, with the onset of industrial society in the 1860s, an increased number of respiratory system diseases, especially bronchus diseases, has been observed due to the burning of fossil fuels/environmental destruction.

Therefore, an effective way to decrease the patient death rate is to reduce the number of patient deaths due to lung disease. To do so, the best way is to avoid the occurrence of lung/bronchus diseases, such as avoiding nicotine exposure and maintaining a healthy lifestyle. However, effective treatment is necessary to diagnose disease in its early stages.

1.2 Types of diagnosis

Early diagnosis can effectively improve the survival rate of patients. According to the literature [1], the five-year relative survival rate has a 20.5% (from 2010 to 2016) benefit from early diagnosis. We classify the existing diagnosis methods of lung cancer into three types: medical images-based, which use medical images (such as Computed Tomography (CT), X-ray, and Magnetic Resonance Imaging (MRI)) for diagnosis; procedure-based, which uses an examination procedure such as bronchoscopy or transbronchial biopsy to diagnosis lesion regions; and biomarker testing-based [9], which uses samples of cells for diagnosis, such as cancer genetic testing and tumor marker testing [10, 11].

1.2.1 Medical images-based diagnosis

With developments in science and technology, more and more imaging techniques are used to obtain medical images for the diagnosis of diseases. There are many kinds of medical imaging methods, such as X-ray (discovered in 1895) [12], CT (invented in 1967) [13], MRI examination (1977/1978, the first MRI scanner) [14], and other imaging techniques. CT images are obtained from a CT machine that records differences in the absorption and transmittance of X-rays of different tissues. CT images allow physicians to see inside the human body without cutting. CT images are processed to find organ regions (lung/bronchus) and lesion regions with segmentation technologies [15] in computer-aided diagnosis (CAD) [16, 17]. The CAD technique originated in the 1950s with the development of modern computers [18]. The CAD system makes it easier to diagnose lung cancer in its earlier stages [19–21].

1.2.2 Procedure-based diagnosis

Procedure-based diagnosis allows physicians to observe a patient's tissues in real-time via devices such as a microscope or bronchoscope [22]. There are many procedures, including microscopy, bronchoscopy, mediastinoscopy and mediastinotomy, and thoracentesis. As an example, in a microscope-based procedure, a microscope is used to inspect the tissues/cells obtained from a suspicious region for further investigation; in a thoracentesis procedure, fluid is removed from the thoracic cavity by needle aspiration for diagnostic and/or therapeutic purposes.

1.2.3 Biomarker testing-based diagnosis

Biomarker testing-based diagnosis uses genetic material such as protein or DNA to diagnosis lung cancer and control the treatment effect. Changes in genetic information such as aggregated proteins determine the therapeutic effect. This type of diagnosis covers three examination types: cancer genetic material examination, PD-L1 examination, and serious injury sign examination.

1.3 Bronchoscopy

1.3.1 Introduction

Bronchoscopy, as an efficient way to diagnose bronchial diseases, has become very popular since the bronchoscope was invented. According to the literature [23], at least 500,000 bronchoscopies were performed each year during the 2000s. There are cur-

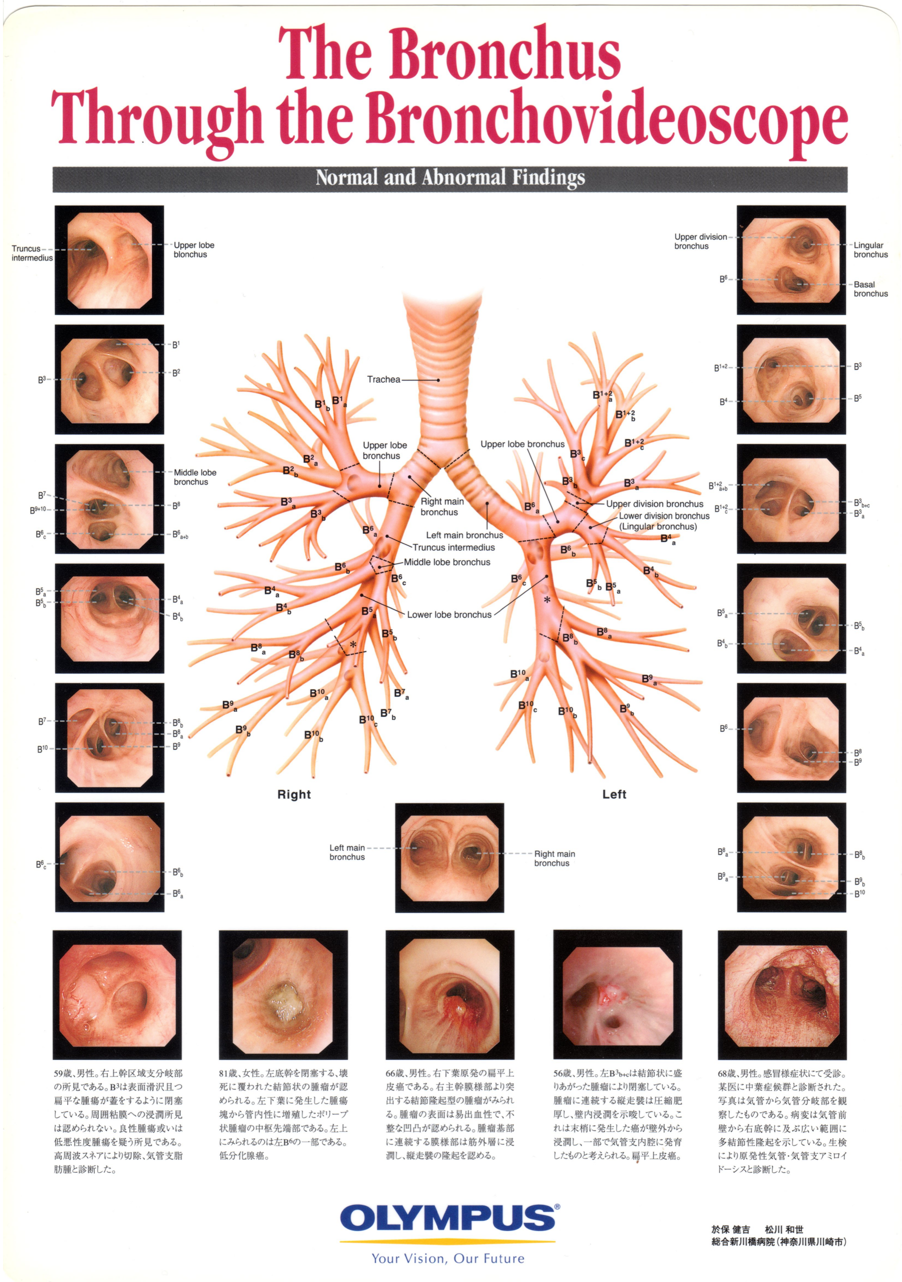


Figure 1.2: Bronchus tree and several RB images obtained in different positions [3].

rently 1342 board-accredited fellows actively performing bronchoscopy in Japan, according to a survey from the Japan Society for Respiratory Endoscopy (JSRE) [24].

Years of experience has made bronchoscopy a safe procedure in the hospitals involved in this survey [24]. In addition to diagnosing lung cancer, there are many other reasons to perform a bronchoscopy: as part of a lung/bronchus examination; identification of a lung infection; to remove a foreign body entering the trachea/main bronchi; and other problems that arise in the lung and bronchus [25].

Before the bronchoscopy procedure, bronchoscopists use preoperative images (e.g., CT images) to diagnosis lesion regions and perform path planning with virtual endoscopy software (VES) [26, 27]. At the beginning of the procedure, the patient will be given a sedative to decrease their breath and heart rates. In addition, medicine is used to numb the patient's throat because, during the examination, the bronchoscope will be inserted into the bronchus via the throat. The patient is kept awake throughout the procedure because the bronchoscopist will need the patient to respond to their instructions. During a bronchoscopy procedure, bronchoscopists insert a bronchoscope (usually a flexible bronchoscope) into the bronchus through the patient's mouth or nose. The patient's bronchial internal information is displayed simultaneously on a monitor. Bronchoscopists manipulate the bronchoscope via an external handle to move to the target area while watching the bronchoscopic video via the monitor [28]. A typical bronchoscopy procedure lasts about half an hour; however, additional time may be needed if a biopsy (e.g., for biomarker testing) is required [25, 29]. Nevertheless, this procedure takes less time for treatment and recovery than other surgeries. Patients will recover and be able to return to work quickly after bronchoscopy.

1.3.2 History of bronchoscopy

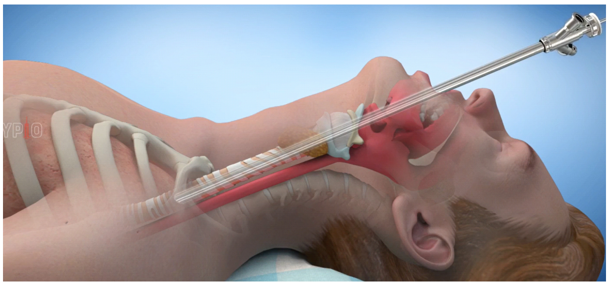
The first bronchial examination appears to be the examination carried out by Gustav Killian in 1876 [30]. He used a rigid endoscope to examine a patient's trachea and main bronchi, and removed a foreign body (pork bone) from the bronchus [31, 32]. With the birth of endoscopes in 1807 [33], the process of bronchial examination has continued to improve. In 1895, Alfred Kirstein reported an operation that used an endoscope to visualize the vocal cords and proximal large airway [32]. There are many types of bronchoscopes, which are classified into rigid (originated in 1897) [32], non-rigid/flexible (originated in 1968) [34], and endobronchial ultrasound-guided transbronchial needle aspiration (EBUS) (originated in 1992) [35–40].

The rigid bronchoscope is a hollow metal tube [41], as shown in the top left of Fig. 1.3. Due to its size and material, its explosive region is limited to the trachea and main bronchus. The flexible bronchoscope is also a hollow tube; however, it is thinner and longer and, most important, its tip can be oriented [42]. Therefore, it can be inserted into the terminal bronchiole. Currently, most examinations use the flexible bronchoscope. The EBUS bronchoscope is an upgraded flexible bronchoscope. Endobronchial ultrasound (EBUS) probes are attached to the bronchoscope's tip. The EBUS bronchoscope is used to sample mediastinal lymph nodes during bronchoscopy.

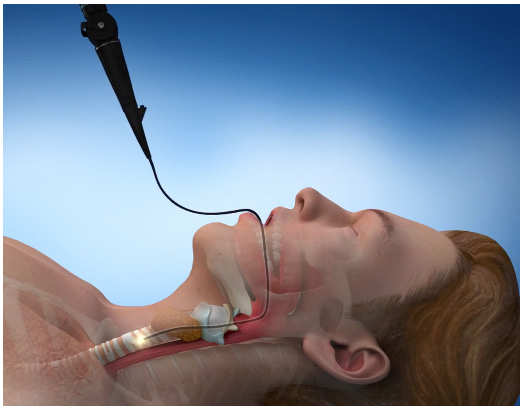
The type of the bronchoscope is selected depending on the purpose of the examination. For example, a rigid bronchoscope is sufficient to remove a foreign object from the bronchus; however, a flexible bronchoscope is necessary for the diagnosis of lesions in the terminal bronchiole. For the sampling of mediastinal lymph nodes, an EBUS bronchoscope is more suitable. Figure 1.3 illustrates examinations using the three types of bronchoscope [4]. The focus of this thesis is the flexible bronchoscope.

1.3.3 Advantages of bronchoscopy

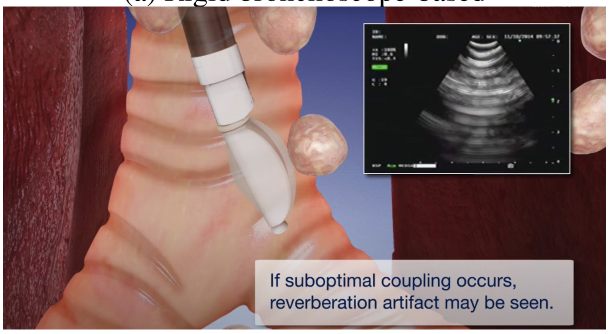
Bronchoscopy is an efficient method for examination; it has many benefits not only for the bronchoscopist, but also for the patients [43–45]. Since the images obtained from this examination type are more comprehensible to the human eye than other types, it is easier for a bronchoscopist to obtain essential information from these images than from



(a) Rigid bronchoscope-based



(b) Flexible bronchoscope-based



(c) Ebus bronchoscope-based

Figure 1.3: Illustration of three types of bronchoscope-based examinations (figures from [4]). (a) shows a rigid bronchoscope-based examination; (b) shows a flexible bronchoscope-based examination; and (c) shows an EBUS bronchoscope-based examination. The figure in the top right is an ultrasound image.

other image types (CT images, X-rays). Moreover, since this examination visualizes the examining regions during the examination, bronchoscopists can appreciate any abnormalities easily in real-time. Abnormal tissue, organ deformation, and even bleeding can be observed through the bronchoscope. Through the benefit of real-time observations, bronchoscopists can perform operations, such as removing copious fluid and foreign objects. Bronchoscopists can even take samples of tissues for follow-up inspection. Patients recover quickly and return to work in a shorter time than with traditional surgery.

1.3.4 Shortcomings of bronchoscopy

However, there are some shortcomings of bronchoscopy based examination, especially for flexible bronchoscope-based examination. Bronchoscopists, especially young physicians, may lose their current field of view (FoV) during the examination, which is caused by the following. First, the insertion cord of the bronchoscope is a soft structure, making it difficult to locate the bronchoscope's tip in the bronchus. Second, the bronchus is a tree-like structure with many branches (as shown in Fig. 1.2). What's more, the bronchial lumen is very similar in appearance, which leads to high similarity in 2D bronchoscopic images. Therefore, it is necessary to use a bronchoscopy navigation system to help the doctor locate the current examination position.

1.4 Bronchoscopy navigation

1.4.1 Introduction

With the popularity of family cars, the term ”navigation” has gradually entered into the public's attention. Navigation originally means a technology that calculates the route from the current location to a destination based on your current location and understanding of the surrounding environment (map) [46]. A driver uses the car navigation system in the following way: firstly, he sets the destination where he want to go in the car navigation system; secondly, the navigation system generates the best route from the current location to the destination based on the existing map (Google map [47–49] or other map [50–52]); thirdly, the navigation system continuously provides direction hints during the journey. The principle of a surgical navigation system is similar to that of a car navigation system. There are several types of navigation systems in clinical application: navigation system for sinus surgery [53–57], navigation system for coronary Artery Stenosis [58, 59], navigation system for knee arthroplasty [60–63],

Table 1.1: Comparison of car navigation and bronchoscopy navigation

Item	Car navigation	Bronchoscopy navigation
Object	Car	Bronchoscope
User	Driver	Bronchoscopist
Map	Google map (Previously downloaded)	Segmented bronchus region
Target	Destination	Lesion areas
Path	Path to destination	Path to lesion
User	Crossroads	Bifurcation
Possbile trouble	Accident and/or other unexpected event	Bubble, blood, etc.

navigation system for abdominal surgery [64–68]. These systems assist the surgeon in different ways in different scenarios. Among them, the bronchoscopy navigation system is used during the examination of lung and bronchus. Physicians are assisted by the preoperative information, the intraoperative information, and the generated navigation information.

To illustrate the similarities and differences between a car navigation system and the bronchoscopy navigation system, we compare the key factors of these two systems in Table 1.1 1.1. To reach the lesion area (destination) smoothly, the navigation system should include the components described in the following subsections.

1.4.2 Components of a navigation system

The bronchoscopy navigation system uses information obtained before and during the surgery to generate navigation information to assist the doctor. Preoperative information generally refers to preoperative images prepossessed by CAD technology. Intraoperative information includes the inference of the endoscope's current position and the navigation path updated in real-time. Therefore, a bronchoscopy navigation system mainly consists of displaying preoperative information, intraoperative position navigation (bronchoscope tracking), and patient-CT registration, integrating preoperative and intraoperative information.

Visualization of preoperative information

Preoperative information mainly refers to the medical images taken before the bronchoscopy. Anatomical structure information is obtained through CAD technology [69–71]. Images obtained before the operation are used as maps of the navigation system. The bronchial area can be obtained by segmentation in CAD to locate the bronchus region and lesion. Bronchial naming technology can obtain the corresponding parts of each branch, and the lesion area and location can be inferred and manually confirmed by a lesion-area recognition algorithm. A volume rendering technique is used to render the CT images into a virtual endoscopic image [72–74]. VES is used to visualize the segmented CT images in most images. A screenshot of VES is shown in Fig. 1.4. In this software, the CT images are viewed in three directions: axial, sagittal, and coronal. The visualized image is shown in the center window. The parameters for volume rendering are shown on the right.

Estimation of camera pose

The estimation of the camera pose is the core of the navigation system. It is essential to obtain the camera position in real-time since the camera position constantly changes during the inspection. The technique of acquiring the camera pose is the so-called bronchoscope tracking. The bronchoscope tracking procedure uses the pre- and intra- operation information to estimate the current position of the camera. We investigate the existing methods for bronchoscope tracking and classify them into four types: (1) manual adjusted tracking (the real bronchoscope is tracked by synchronize the virtual bronchoscope view according to the real bronchoscope view); (2) video-CT image registration-based tracking (the camera pose is estimated by maximizing the similarity between the virtual bronchoscope view and the real bronchoscope view iteratively); (3) additional sensor-based tracking (the camera pose is estimated from the output of an

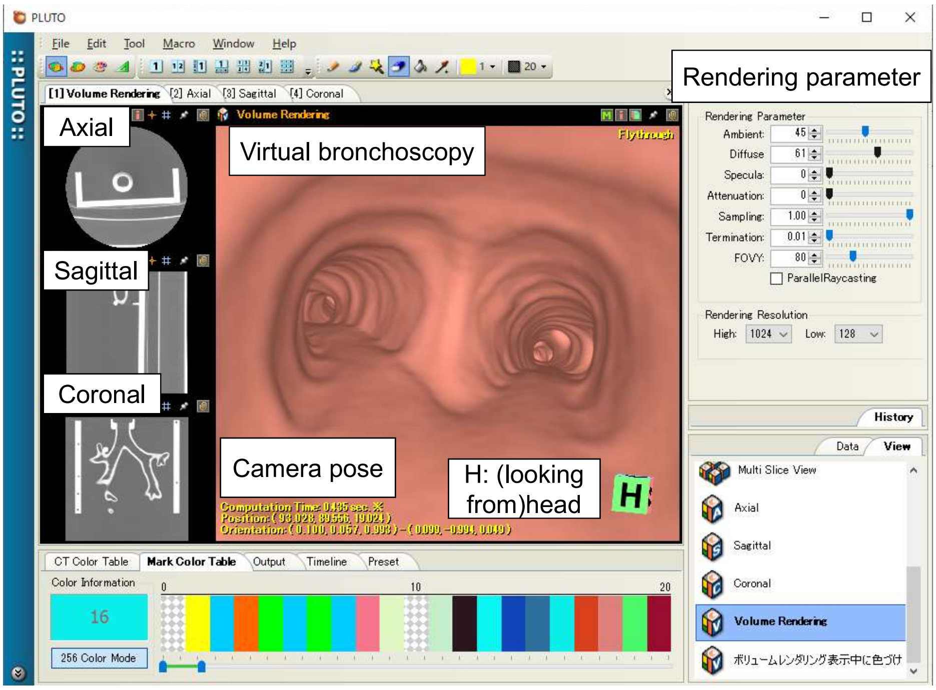


Figure 1.4: Screenshot of a virtual endoscopy software. This figure shows the CT images in axial, sagittal and coronal cross sections, the main interface shows rendered CT volume; the camera pose is shows in yellow text in the left bottom; the body direction is shown in right bottom.

additional sensor); and (4) deep learning-based tracking (the camera pose is estimated by using massive paired image-pose data to train a deep learning model). We introduce these methods in detail in the following chapters.

Patient-CT registration

Since the preoperative CT image coordinate system is different from the intraoperative camera coordinate system, it is necessary to use the common information (such as landmarks or anatomical structures) in the two coordinate systems for registration. The camera position and surgical tools during the operation will be displayed in the navigation system using the registration results. Therefore, the accuracy of the registration results will directly determine the visualization of the navigation information. Generally, the registration step is needed to be performed once before the navigation starts. However, if the camera fails to track during the operation or the tracking error gets large, the registration step may be performed again. We investigate the existing registration methods in navigation systems and classify them into three types: (1) landmark-based (the registration is carried out using anatomical landmarks or artificial markers for registration [75]); (2) centerline-based [76, 77] (the registration is carried out by using the medial line extracted in preoperative CT images and the trajectory of the bronchoscope camera); and (3) bronchoscope tracking-based [78] (the registration is carried out simultaneously with the bronchoscope tracking procedure. Notably, video-CT-based bronchoscope tracking achieves patient-CT registration and estimates the camera pose simultaneously).

Other auxiliary functions

Several auxiliary functions are integrated with the main navigation function to provide better support to physicians. For example, the bronchial deformation exists during bronchoscopy and it will cause the difference between preoperative and intraoperative information. If a navigation system implements this function, the pre- and intra- operation error can be reduced by simulating the bronchial deformation [79, 80]. Even though more functions will make a complex navigation system, it is beneficial to the bronchoscopist.

Table 1.2: Types of bronchoscopy navigation systems

System name	Tracking method	Patient-CT registration	Description
SuperDimension [75, 82]	EM sensor-based	Landmark (plastic sheath)-based registration	Wide clinical application
LungPoint [83]	Manual tracking	Image-based registration	Generates path to target, Clinical application
Veran SPiNDrive [84]	EM sensor-based	Not mentioned	Lack of technique details and clinical application
Bf-NAVI [85]	Manual tracking	Image-based registration	Virtual bronchoscopy navigation for EBUS

1.4.3 Bronchoscopy navigation systems

The first bronchoscopy navigation system (BNS) [81] was developed in 1998, since then, many solutions have been developed. Among these solutions, several BNSs have been used clinically and most are still being researched. We conducted a survey to investigate current navigation systems, including commercially available systems and those under development. Several commercial platforms are shown in Table 1.2. There are many navigation schemes under development. Since the performance of navigation systems is mainly based on the results of bronchoscope tracking, we investigated the existing bronchoscope tracking methods. In the following section, we describe the existing bronchoscope tracking methods in detail.

1.5 Bronchoscope tracking

Bronchoscope tracking manages to estimate the camera pose of the bronchoscope during the examination. In addition to the tracking methods shown in Table 1.2, there are many tracking methods under development. In this Section, we describe the state-of-the-art bronchoscope tracking methods in detail.

Existing bronchoscope tracking methods can be roughly classified into four types: video-CT-based tracking [86–89], additional sensor-based (electromagnetic (EM) sensor) tracking [90–93], hybrids of two methods [91, 94, 95], and deep learning-based tracking.

Pose from video-CT registration

Video-CT-based bronchoscope tracking uses RB images and virtual bronchoscopic (VB) images for camera pose estimation. RB images are color images obtained in real-time from a bronchoscope; VB images are images rendered from preoperative CT images. The camera pose is estimated by maximizing the image similarity between RB images and VB images [78, 81, 95–97]. Deguchi et al. improved the image registration procedure by using selected subblocks instead of the entire RB and VB images [86]. Subblocks are selected according to the characteristic structures (folds, patterns, bifurcations, and so on). Tracking performance is greatly improved in images containing rich characteristic structures and is weak in textureless images. Luo et al. took more structural information (such as luminance, the contrast between selected blocks) into consideration while calculating the image similarity [89, 91]. Tracking accuracy decreased from 14.6 mm to 4.5 mm, and processing can be achieved in real-time using a GPU for acceleration.

In addition to the improvement of the similarity function, Shen et al. improved the conventional video-CT-based method by using depth information for image registration [98, 99]. The corresponding depth image of RB images is estimated using shape from shading (SFS) [100], while the corresponding depth images of VB images are the projection of CT images on a virtual bronchoscope. The camera pose is estimated by finding the maximum similarity between the two types of depth images. The depth image-based registration decreased the influence of illumination artifacts.

Pose from additional sensor

Sensor-based bronchoscope tracking uses the output of an EM sensor that is integrated onto the camera tip of a bronchoscope for navigation. There are three ways to use the sensor's output: (1) transform the 3D output of the sensor from sensor coordinates into real camera coordinates by using the result of a hand-eye calibration procedure before the examination [101, 102]; (2) transform the 3D output directly into the coordinates of CT images to generate virtual bronchoscopic images for navigation [103–105]; and (3) hybrid tracking using the output of the sensor and the video-CT-based tracking [87, 94, 106, 107]. Merritt et al. improved the similarity calculation procedure and the procedure for generating virtual bronchoscopy, which increased bronchoscope tracking accuracy while decreasing computation time [95].

Pose from deep learning

With the improvement of hardware and deep learning algorithms, deep learning has become popular in recent years. Some research groups try to apply deep learning algorithms to bronchoscope tracking [108–110] and even bronchoscopy navigation [111].

Shen et al. used deep images generated by deep learning to improve video-CT-based tracking [108]. They changed the image matching process from the original color-based image domain to the depth domain to improve the robustness of bronchoscopy tracking. The depth image corresponding to the color image is generated by the pre-trained model; the depth image corresponding to the CT image is calculated using the depth image definition. The image matching process is based on two depth images. The camera pose is the pose of the virtual camera when the two depth images get the maximum similarity.

Sganga et al. used a localization network based on a convolutional neural network (CNN) to achieve bronchoscopy tracking [109]. The essence of this method also uses the idea of video-CT-based tracking. The similarity between the virtual image and the real image is calculated to reduce the position difference between the virtual camera and the real camera. The camera pose is estimated by minimizing the difference between the virtual camera and the real camera.

There are many applications of deep learning in bronchoscopy. McTaggart et al. used a classification network to classify the RB images into informative or uninformative [111]. One image is judged as informative or uninformative according to its imaging quality. The experimental results showed that the deep learning-based method is superior to the traditional edge-based method.

1.6 Positioning of this study

This section shows the positioning of our study. As it is described in the previous Chapter, the bronchoscopy navigation technique acts as an additional important method to assist physicians during bronchoscopy. Bronchoscopy navigation is one of the most important components in computer-aided surgery (CAS), its performance such as accuracy and robustness are vital to a patient's life. Common bronchoscopy navigation systems rely on the result of the bronchoscope tracking and the anatomical structure of the bronchus. Therefore, this thesis mainly researched the bronchoscope tracking and the bronchial anatomical structure extraction, which are listed in the following:

(1) RB image-based precise bronchoscope tracking for bronchoscopy navigation

(2) RB image-based bronchial orifice (BO) segmentation for scene understanding

(3) BO changes-based coarse bronchoscopy navigation

In the following section, we will introduce each topic in details by considering the previous works.

(1) RB image-based precise bronchoscope tracking for bronchoscopy navigation

Bronchoscope tracking is an essential part in a bronchoscopy navigation system. The performance of bronchoscope tracking decides the performance of bronchoscopy navigation. The previous video-CT-based bronchoscope tracking depends mainly on the similarity of image structures between RB and VB images or additional sensor-based tracking. However, for video-CT-based tracking, the difference between preoperative and intraoperative images decreases the accuracy of bronchoscope tracking. The organ deformation and bubbles that exist in RB images are caused from patient breathing, which do not exist in VB images. These differences may lead lead to the increment of the tracking error, and so as the navigation. Moreover, registration error increases in the region lacking local texture information (such as bifurcations or folds). What's more, video-CT-based methods require a huge amount of computation time for VB images generation and images registration, which makes it difficult to perform navigation in real-time for most navigation applications. On the other hand, EM sensor-based tracking also have shortcomings [86, 87] even though there are some clinical applications. The positional data obtained from EM sensor may have jitters, which results to unstable navigation results; a sensor may also be affected by metallic surgical tools in the operating room; and a bronchoscope equipped with an EM sensor cannot be inserted into terminal bronchi due to its excessive size [86, 87]. Therefore, We use the pure RB images for bronchoscope tracking to decrease the difference between preoperative and intraoperative. RB images are processed by camera tracking methods in computer vision. Experimental results showed that it is feasible to use the proposed method for bronchoscope tracking.

(2) RB image-based BO segmentation for scene understanding

Scene understanding is an important task in surgeries. It contains many sub-taskssuch as image classification, textual annotation, and object segmentation [112]. Due to the complexity of the bronchus scenes, object segmentation in the bronchus scenes, such as the segmentation of the anatomical structure (the BO and the carina) and surgical tools is more meaningful. The BO region is one of the representative characteristics in the bronchus scene which is important in both navigation bronchoscopy and surgery assessment. Previous image appearance-based work used image appearance and the gradation of the RB image to segment the orifice region, which behaved poorly in complex scenes including bubbles or changes in illumination. Therefore, we propose a depth image-based segmentation algorithm to obtain a better segmentation result of BO even in the complex scenes. Experimental results show that the BO region is accurately segmented in RB images.

(3) BO changes-based coarse bronchoscopy navigation

Conventional navigation systems based on the result of precise bronchoscope tracking, ex. video-CT-based and EM sensor-based bronchoscope tracking, in which the tracking error is accumulated and lead to the failed of navigation. To overcome the conventional precise bronchoscope tracking-based navigation systems, we propose a coarse navigation system which is based on the result of coarse bronchoscope tracking. This bronchoscope tracking operates by using the BO region's changes to estimate branching level. Preliminary experimental results showed that the branching level is correctly estimated. To decrease the tracking errors accumulated during bronchoscope tracking, a coarse navigation scheme is used to identify the branches in which a bronchoscope is currently observed. Therefore, a branching level estimation algorithm based on the changes of the extracted anatomical structure is proposed. Experimental results showed that the accuracy of the proposed branching level estimation methods is 87.6 %.

1.7 Structure of this thesis

This thesis consists of five chapters. Chapter 1 describes the backgrounds of this research, the definition and the history of navigation bronchoscopy, and the positioning of this study.

Chapter 2 describes a pure RB image-based method (visual SLAM) for precise bronchoscope tracking. We show the mechanism of visual SLAM, our effects to make it suitable for bronchoscope tracking and the performance of improved visual SLAM in bronchoscope tracking in this section. The root mean square error (RMSE) value of the estimated camera pose was 3.02 mm while that of the original was 3.61 mm.

Chapter 3 describes a depth image-based BO segmentation method. We introduce the proposed method from the preparation of the training dataset and the decision of the BO region in depth images. This method is validated in several *in-vivo* videos. We manually found BO regions as ground truth to evaluate the BO segmentation method. The performance of the proposed method was shown in this Chapter.

Chapter 4 describes a new coarse bronchoscope tracking scheme by using the changes in anatomical structure: BO to estimate the branching level. The number of the BO is counted from the depth image, which is estimated by the deep learning method described in Chapter 4. Meanwhile, the movement of the bronchoscope camera is estimated from feature-based camera motion estimation. The branching level is estimated from the aforementioned method. Experimental results showed the performance of the proposed branching level estimation method.

Chapter 5 describes a comparison of these two tracking methods and the future work of this research.

Chapter 2

Computer vision-based precise bronchoscope tracking

This chapter describes the author's work on computer vision-based bronchoscope tracking. Due to the complex anatomical structure of bronchi and the resembling inner surfaces of airway lumina, bronchoscopic examinations require additional 3D navigational information to assist the physicians. A bronchoscopy navigation system provides the position of the endoscope in CT images with augmented anatomical information. To overcome the shortcomings of previous navigation systems, we propose using a technique known as visual simultaneous localization and mapping (SLAM) to improve bronchoscope tracking in navigation systems. This chapter explains the proposed method from three aspects: (1) image-guided bronchoscope tracking; (2) proposed method for bronchoscope tracking; and (3) the performance of the proposed method.

2.1 Background

The bronchoscopy navigation systems mainly contain the bronchoscope tracking part, CT-patient registration part, and the visualization part. The structure of the navigation system is shown in Fig. 2.1. As explained in the first Chapter, the bronchoscope tracking part manages to estimate the bronchoscope's camera pose. The patient-CT registration part manages to find the coordinate relationship between preoperative CT images and the intraoperative RB images. And the visualization part shows the integrated preoperative and intraoperative information.

Conventional video-CT-based bronchoscope tracking is influenced by the difference between two image types. Therefore, a tracking method that uses the pure bronchoscope image may benefit the bronchoscope tracking procedure.

2.2 Introduction of computer vision

Computer vision is a field that lets computers know the world from digital video or images[113–115]. It operates by extracting the useful information from the inputs (video/images) and processing it according to the demonstrated theorem. It is a way to quantify the world, analyze the world, and interpret the world from the digitized 'eye'. Therefore, several methods in the computer vision field are used to estimate the bronchoscope pose by using the RB images as input.

2.2.1 Methods to recover 3D from 2D

The 2D image doesn't contain the 3D information completely, however, the camera pose is expected to know in 3D. Therefore, the obtained 2D images are used to recover 3D information.

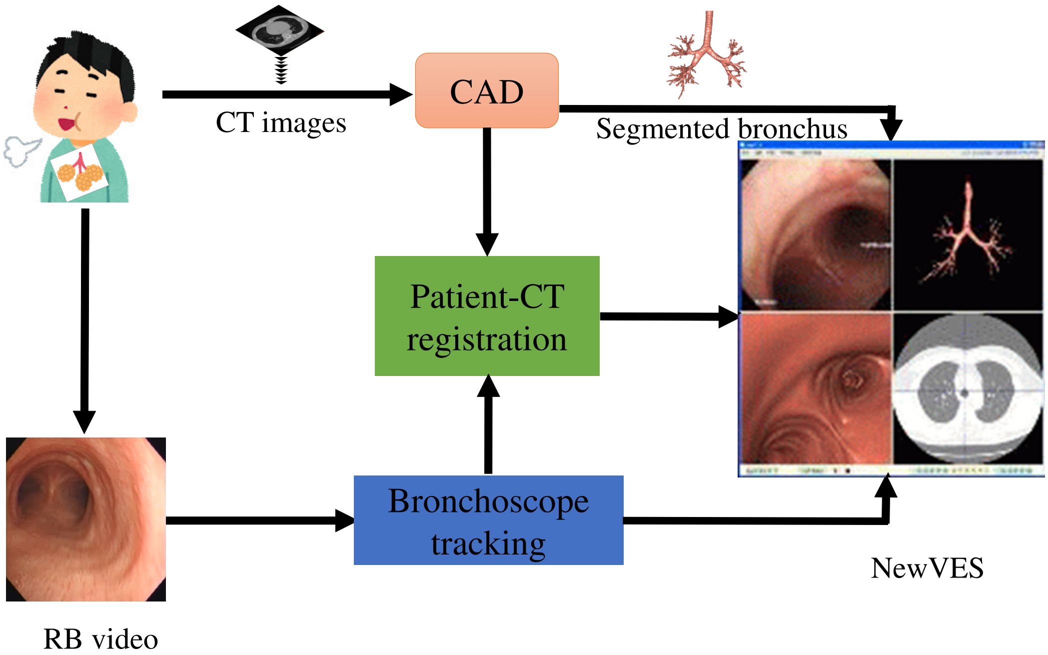


Figure 2.1: System structure of navigation system: preoperative CT images are used to segment the bronchus, which is used as the input of the virtual bronchoscopy software and patient-CT registration; the RB video is used to estimate the camera pose by using bronchoscope tracking (visual SLAM). The output of bronchoscope tracking are the camera pose and bronchus shape. The bronchus shape are used for patient-CT registration together with the shapes from CT images. Camera pose is transformed using the patient-CT registration result.

There are many methods to recover the 3D camera pose from 2D images. For example, the structure from motion (SfM) technique is used to estimate the 3D positional information by using 2D images as input. This method assumes that there are enough shared views among these images. The 3D information is estimated by solving a function that minimizes the reprojection error of found 3D-2D matches. Recently, deep learning-based methods train a network to find the relationship in prepared data. For example, Visentini-Scarzanella et al. train a dual CNN network to continuously translate the RB images into VB images and then translate them into the corresponding depth images [116].

2.2.2 Visual SLAM

Simultaneous localization and mapping (SLAM) is an approach to estimate the senor pose and the environment from the sensor data. There are many types of SLAM scheme according to the sensor used: Lidar, single camera (monocular), multi-camera (stereo camera or RGB-D). The SLAM schemes that process visual information from camera/cameras are defined as visual SLAM [117].

Through two decades of development, many excellent solutions have been proposed since the first SLAM application appeared in 2003 [118]. These solutions have been tested in minimally invasive surgery (MIS) [119–123]. However, the size of the bronchus, especially the terminal bronchioles, limit the size of the bronchoscope. In most scenes, only the size of an optical camera is allowed at the tip of the bronchoscope. To examine the terminal branches as much as possible, the bronchoscope tip should be as small as possible, therefore, a small monocular camera is used. At the beginning of the study, a survey is carried to investigate the conventional visual SLAM solutions, including their advantages and disadvantages, which are shown in Table 2.1. Many solutions are aiming at the monocular camera, as shown in Table 2.1. In recent years, deep learning-based methods have also become popular and are applied into SLAM approaches [46, 133–144]. Visual SLAM is originally designed to estimate camera poses and reconstruct 3D environments around a camera by using video of real-world scenes. However, bronchus scenes are different from ordinary scenes. Most real-world scenes are static with rich textures, while bronchus scenes are textureless with complex tree structures. The difference between each pixel in one frame is small, as is the difference between adjacent frames. Such homogeneous scenes result in additional mismatches during camera tracking and 3D reconstruction. Moreover, the illumination in bronchus scenes changes strongly. Since the dense SLAM [126, 128] uses image pixels directly, which may either be influenced by illumination or be computationally demanding. A

Table 2.1: SLAM solutions and a introduction of the details

Name of scheme	Implementation method	Usage scenarios	Memo
MonoSLAM [124]	Feature point-based Extend Kalman filter	Indoors (Kitchen)	First visual SLAM in real-time
PTAM [125]	Feature point-based Optimization	Indoors (Desk)	For AR application
DTAM [126]	Dense Optimization	Indoors (Desk)	Run on GPU
SVO [127]	Semi-direct	Micro Aerial Vehicles	Use feature point and/or whole pixel
LSD-SLAM [128]	Direct based Optimization	Outdoor (Road)	Semi dense
DSO [129]	Feature point-based Optimization	Indoors and outdoors	Best direct-based
ORB-SLAM [130–132]	Feature point-based Optimization	Indoors and outdoors	Best feature point-based

feature-point-based SLAM scheme, monocular ORB-SLAM [130] is selected, for bronchoscope tracking among the existing visual SLAM frameworks.

2.3 Purpose of this Chapter

In this Chapter, a new bronchoscope tracking method is proposed by using the visual SLAM technique in the computer vision. Visual SLAM uses pure bronchoscopic images as input to estimate the camera pose and the environment surrounding the camera. The outputs of the visual SLAM are used for patient-CT registration and the virtual bronchoscopy software. The structure of the system is shown in Fig. 2.1.

2.4 Proposed method for bronchoscope tracking

This Section explains the proposed method for bronchoscope tracking. Since the original visual SLAM is designed for ordinary scenes, it is essential to change before applying it to the bronchus scene. The details of the proposed method are described in Section 2.4.2.

The scheme structure of visual SLAM-based navigation is shown in Fig. 2.2. In the following section, the proposed tracking procedure and its evaluation are introduced in three parts: (1) details of the original visual SLAM technique; (2) improvement of local feature matching in coarse-to-fine tracking; and (3) quantitative evaluation of the camera trajectory.

2.4.1 Visual SLAM for bronchoscope tracking

ORB-SLAM is a semi-dense tracking solution that extracts **O**riented FAST and **R**otated **B**RIEF (ORB) [145] features in images for camera tracking and 3D reconstruction. ORB is a method for keypoint detection and description that outperforms other methods in processing time while keeping a good accuracy [145]. ORB-SLAM uses many successful theory in previous literature, such as the usage of the bundle adjustment for the optimization of the reprojection error[146, 147]; the usage of bag of words (BoW) for the recognition of the visited place [148, 149]; and so on. ORB-SLAM is introduced from three aspects: initialization of RB image coordinates; pose estimation of each frame; and creation of new 3D points.

Initialization of RB image coordinate

To determine the coordinate system of a real bronchoscope, an initialization procedure is used prior to tracking. This procedure estimates an initial relative pose between two adjacent frames and finds an initial set of 3D points. As shown in Fig. 2.3, two adjacent frames are selected to find the 2D-2D feature matches that are used to calculate relative pose between two frames. The initialization procedure runs iteratively until a motion

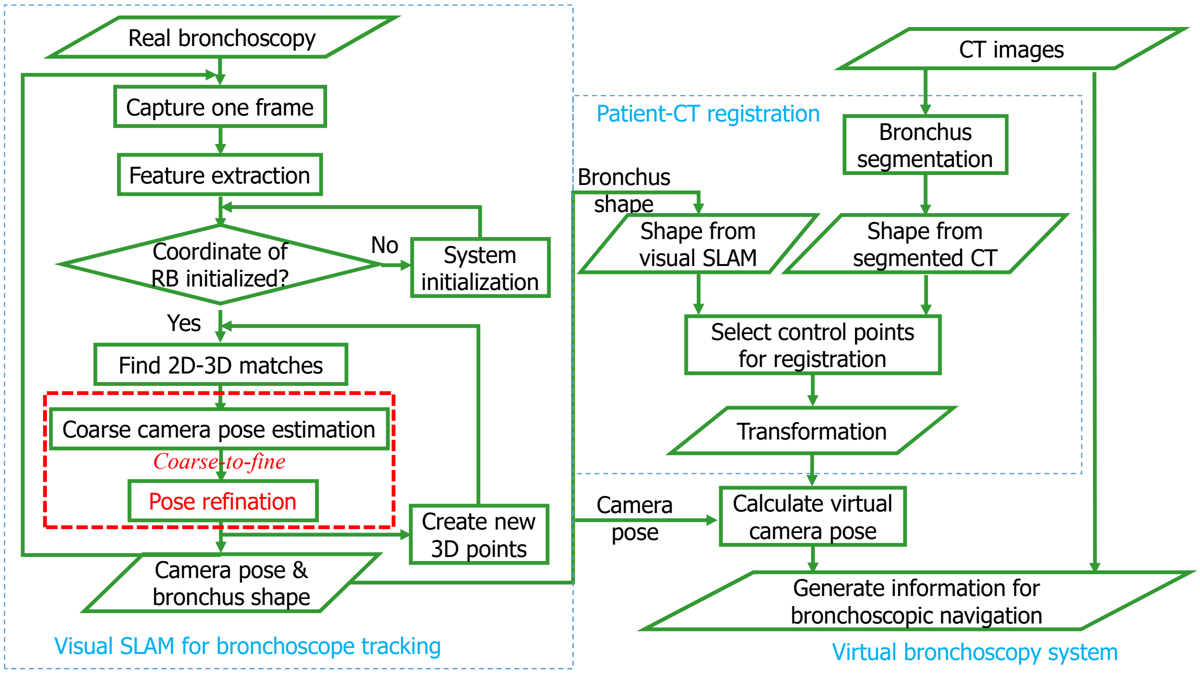


Figure 2.2: System structure of improved visual SLAM and its application to bronchoscopy navigation. Visual SLAM is used as the bronchoscope tracking method; the output of visual SLAM and processed CT images are used for patient-CT registration; and the registration results together with the output of visual SLAM are used to transform the camera pose from RB to VB to generate navigational information.

hypothesis that satisfies most 2D-2D matches is found.

Under this hypothesis, the origin of real bronchoscope coordinates is set as the center of the first frame (Fig. 2.3). The reconstructed 3D points represent an initial estimation of the bronchus lumen (the map). More details about the initialization of real bronchoscope coordinates can be found in Fig. 2.3.

Finally, an estimate of the camera pose **P** and the bronchus lumen are obtained after initialization. The camera pose has 6 degrees of freedom, which has a format of

$\begin{bmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0} & 1 \end{bmatrix}$, where $\mathbf{t} = [x \quad y \quad z]^\top$ is the 3D camera position; and $\mathbf{R}$ is the camera orientation in a rotation matrix. The bronchus lumen is represented by $\mathbf{B} = \{\mathbf{b}_i\}$, where $\mathbf{b}_i$ is the i-th 3D coordinate of reconstructed points. The camera position $\mathbf{t}$ and a 3D point $\mathbf{b}_i$ have lost the scale information due to the normalization during initialization. However, the scale information is recovered to compute the transformation from real bronchoscope coordinates to the CT coordinate system via the patient-CT registration procedure as described in Section 2.4.3.

Estimation of camera pose

The tracking procedure uses the reconstructed 3D points together with the feature points in one frame for the estimation of the camera pose. This uses a coarse-to-fine procedure: during the coarse portion of the procedure, a subset of reconstructed 3D points is selected to find the corresponding 2D points in a frame, and the camera pose is estimated using the found 3D-2D matches; the fine portion of the procedure uses the 3D points nearby to find additional 3D-2D matches to refine the obtained camera pose [130].

During the coarse portion of the procedure, the found 3D-2D correspondences are used to estimate the camera pose at the k-th frame $\mathbf{F}^{(k)}$ using the following equation:

$$^{S}\mathbf{P}^{(k)} = \arg\min_{\mathbf{P}} \sum_{j} \rho \left(\left\| \mathbf{u}_j^{(k)} - f(\mathbf{P}, \mathbf{b}_j) \right\| \right), \tag{2.1}$$

where ρ is the Huber influence function [130], and $\mathbf{b}_j \leftrightarrow \mathbf{u}_j^{(k)}$ is the j-th 3D-2D match. Function f is the projection function, which projects 3D point $\mathbf{b}_j$ onto the k-th frame using camera pose $\mathbf{P}$. This equation finds the best camera pose that minimizes the re-projection error on the k-th frame, and a set of 3D-2D matches (inlier matches) support

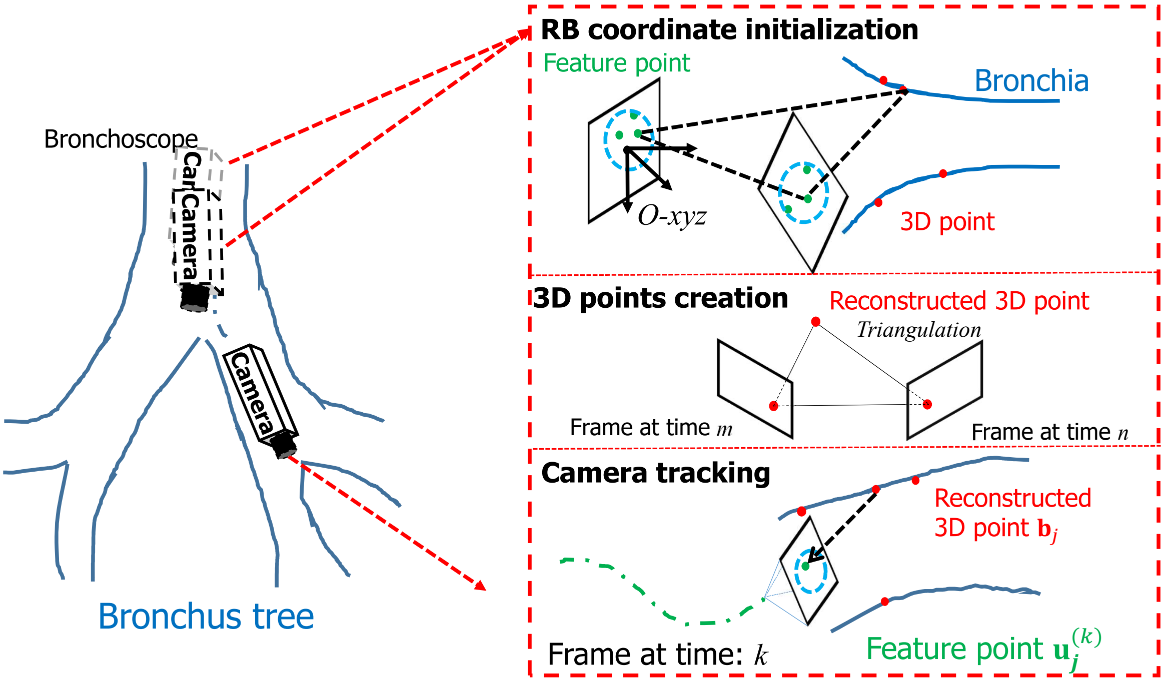


Figure 2.3: Illustration of three-part procedure of visual SLAM. Initialization uses 2D-2D matches in two frames for the initialization of real bronchoscopy coordinates; camera tracking uses 3D-2D correspondence for estimation of the camera pose; and 3D point creation uses the previous two frames to find new 3D points.

this pose. The pose of the k-th frame is denoted as ${}^{S}\mathbf{P}^{(k)}$.

After the camera pose is estimated in the coarse portion of procedure, a pose refinement procedure is used to find more 3D-2D matches. The details of this procedure are introduced in Section 2.4.2.

Creation of new 3D points

For the *p*-th frame, assuming that its camera pose has been estimated, a set of candidate frames are selected for the creation of new 3D points from previous frames according to the number of shared 3D points. Each feature point is checked if its 3D position has been recovered. For the unreconstructed feature points, the corresponding points in other frames are matched using epipolar geometry. The 3D position of one match is computed using

$$\mathbf{b} = triangulate(\mathbf{u}_j^{(p)}, \mathbf{u}_j^{(q)}), for\ each\ \mathbf{u}_j^{(p)} \leftrightarrow \mathbf{u}_j^{(q)} \tag{2.2}$$

where *triangulate* is the triangulation function that recovers the 3D position of 2D-2D matches, $\mathbf{F}^{(q)}$ satisfies $\left\{\mathbf{F}^{(q)} \mid \mathbf{F}^{(p)} \cap \mathbf{F}^{(q)} < threshold\right\}$, and $threshold$ is the number of shared 3D points. Each correspondence is triangulated to find its 3D position, and this 3D point is validated against narrow criteria (reprojection error, parallax of two frames, etc.) before it is used in the coming tracking procedure.

2.4.2 Improved pose refinement for bronchoscope tracking

The pose refinement procedure uses the 3D points near frame $\mathbf{F}^{(k)}$ to refine the pose obtained in equation 2.1. The refinement procedure contains three main steps: (1) selection of candidate frames based on shared 3D points among the frames; (2) increment of 3D-2D matches in the candidate frames; and (3) final pose optimization using the found 3D-2D matches. Due to the homogeneous bronchus scene, there are many mismatched 3D-2D correspondences. Therefore, the search region of the 3D points are limited to the front of the camera, which is achieved by using camera posture as a factor as a limitation while the procedure finds 3D-2D matches—which is called posture guided feature matching [150].

Candidate frame selection

This procedure selects the candidate frames used for finding 3D-2D matches. Frames near $\mathbf{F}^{(k)}$ are selected according to the shared 3D points in the previous tracking procedure. One frame is judged as a candidate frame only if it has enough shared 3D points with $\mathbf{F}^{(k)}$ [130].

Increment of 3D-2D matches

3D points located in the selected candidate frames are collected and used to find their corresponding points in $\mathbf{F}^{(k)}$. To make the tracking procedure more suitable for bronchus scenes, the 3D points in front of the camera at the moment (i.e., those that are visible) are selected instead of using all 3D points. For a 3D point **b**, the corresponding point on $\mathbf{F}^{(k)}$ is searched in a region $\pi(\mathbf{b},{}^{S}\mathbf{P}^{(k)}) \pm r$, where r is the number of pixels. The original visual SLAM uses the distance between the descriptors to make the decision, which may result in mismatches in homogeneous bronchus scenes. The corresponding point of **b** in the nearest frame to $\mathbf{F}^{(k)}$ are used. One point should also satisfy $\left\|\pi(\mathbf{b},{}^{S}\mathbf{P}^{(k)}) - \pi(\mathbf{b},{}^{S}\mathbf{P}^{(l)})\right\| < d$ before it is considered as the corresponding point, where d is a pre-set threshold of distance, and $\mathbf{F}^{(l)}$ is the nearest candidate frame to $\mathbf{F}^{(k)}$. This is a stricter criterion than in the original ORB-SLAM. Therefore, more accurate 3D-2D matches than the original visual SLAM are found.

Local bundle adjustment

With the found 3D-2D matches, a bundle adjustment procedure is used to find the minimum reprojection error. This procedure is similar to the pose estimation procedure described in Section 2.4.1, however, this procedure contains more 3D-2D matches that help to refine the camera pose.

2.4.3 Quantitative evaluation of tracking

For the quantitative evaluation of the camera pose, an investigation is carried to investigate the existing evaluation schemes. A general evaluation method uses data from an additional sensor, however, sensor-based evaluation may suffer from electromagnetic jitter and the much effect are needed on finding corresponding timestamps. Considering the scale to be ambiguous in visual SLAM results, the output of visual SLAM are evaluated by using the bronchus shape information from CT images. The validation technique contains three steps: (1) the creation of a ground truth by the comparison of RB and VB; (2) transformation of the camera pose from real bronchoscope coordinates to CT coordinates by using the results of patient-CT registration; and (3) calculation of the root mean square error (RMSE) value between the transformed camera pose and ground truth data.

Creation of ground truth

Several frames whose poses have been successfully estimated by visual SLAM are picked and their camera pose in CT coordinates are saved as ground truth. To create the ground truth of one frame, the pose of the virtual camera is updated manually to generate the virtual bronchoscopic images that is most similar to this frame. The software described by Nimura [151] is used in experiment to create virtual bronchoscopic images. The pose of the virtual camera in the CT coordinates is selected as the ground truth of the camera pose. For the sake of fairness, the ground truth data is checked by two engineers and a physician. Four ground truth sets of camera poses are obtained.

Estimation of transformation

For the validation of camera poses, the camera pose is converted from visual SLAM to CT coordinates. This procedure starts from the transformation estimation of two coordinates. CT coordinate is represented by a segmented bronchus shape. The bronchus region in CT images are segmented using CAD technique; then the points located on inner surface of the bronchus are exported as the bronchus shape. This procedure is illustrated in Fig. 2.4. The shape from visual SLAM is used to find the corresponding control points. The control points are selected manually in a bronchus shape from visual SLAM and CT images. Then these point-pairs are used to estimate a 7 degrees of freedom transformation. This transformation between two shapes is estimated by using

$$^{C}\mathbf{T}_{S} = \arg\min_{\mathbf{T}} \frac{1}{N} \sum_{i=1}^{N} \left\| \mathbf{p}_{C}^{(i)} - \mathbf{T}\mathbf{p}_{S}^{(i)} \right\|, \tag{2.3}$$

where $\mathbf{p}_{C}^{(i)}$ is the i-th selected point in bronchus shape from segmented CT images, $\mathbf{p}_{S}^{(i)}$ is the corresponding point of $\mathbf{p}_{C}^{(i)}$ in shape from visual SLAM. $^{C}\mathbf{T}_{S}$ has a format of $s\begin{bmatrix}\mathbf{R} & \mathbf{t}\\ \mathbf{0} & 1\end{bmatrix}$, which contains the scale s, rotation $\mathbf{R}$, and translation $\mathbf{t}$. At least four pairs ($N >= 4$) are needed to find the transformation.

Evaluation of tracking accuracy

The estimated camera pose (position) is evaluated by calculating the RMSE value between the ground truth and the transformed camera pose:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N} \left\| \mathbf{q}_{gt}^{(i)} -^{C} \mathbf{T}_{S}\mathbf{t}_{S}^{(i)} \right\|^{2}}, \tag{2.4}$$

where $\mathbf{q}_{gt}^{(i)}$ is the i-th ground truth of the camera pose, $\mathbf{t}_{S}^{(i)}$ is the i-th camera position from visual SLAM. Since the ground truth of the camera trajectory are created using the selected RB images, there is no need to consider the aforementioned problems that influence the evaluation results. Selected frames are well-distributed in the camera

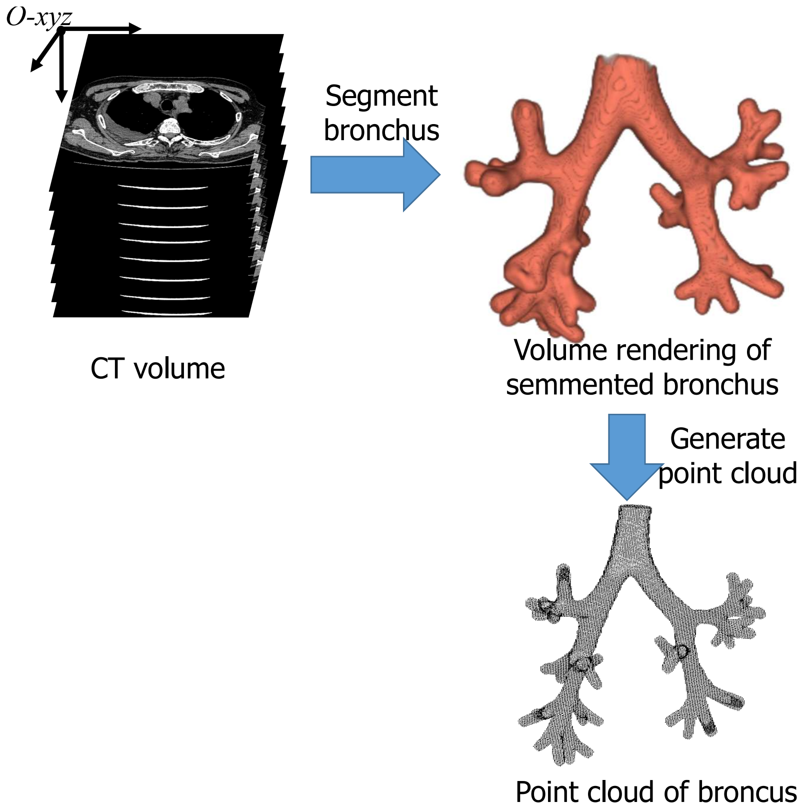


Figure 2.4: Flowchart of Generation of bronchus shape from CT images

trajectory.

2.5 Bronchoscopic video and CT images

Two rubber phantoms are used for evaluation of the proposed method. The videos for processing are captured by using a bronchoscope (BF-200, Olympus, Tokyo) to explore the bronchus phantoms. These videos are at 30 fps with a resolution of 444×440 pixels. These videos last from 1 to 4 minutes.

CT images of the rubber phantoms are taken using a helical CT scanner. The acquisition parameters of these CT images are shown in Table 2.2.

Table 2.2: The acquisition parameters of CT images

Case	Slice size (pixels)	Pixel size (mm)	Number of slices	Thickness (mm)	Reconstruction pitch (mm)
Case 1	512×512	0.39×0.39	667	0.50	0.30
Case 2	512×512	0.63×0.63	709	1.00	0.50

(Kernel function is FC13 and the CT scaner type is Aquilion.
Aquilion: Aquillion 16, Toshiba Medical Systems Inc., Tokyo)

Table 2.3: Information of the ground truth

Trial	Case	Total frames	GT's FN	Description
1	Case 1	3657	42	Static:TR$\rightarrow$ LMB
2	Case 1	5372	56	Static:TR$\rightarrow$ LMB$\rightarrow$ TR$\rightarrow$ RMB
3	Case 1	3313	27	Large deformation:TR$\rightarrow$ LMB
4	Case 1	3111	22	Slight deformation:TR$\rightarrow$ LMB
Average	-	3863	37	-

GT's FN: number of frames with ground truth; TR: trachea; LMB: left main bronchus; RMB: right main bronchus.

2.6 Experimental setup

A conventional PC (3.00 GHz Intel Xeon processor, GTX 750 Ti graphics card, Ubuntu 16.04 operation system) was used to process the video sequence. 8 *ex-vivo* videos were recorded by exploring two rubber bronchus phantoms with a bronchoscope (shown in Fig. 2.5), and two of them had simulation of breath (deformation) from touching the outside of the phantom. The frequency of breath in the third trial was about 0.4 per second and in fourth trial was about 0.3 per second. Illustration of the simulation procedure was shown in Fig. 2.6. Four trials were selected to make the ground truth of the camera trajectory for quantity evaluation. Detailed information about the videos is shown in Table 2.3. The bronchoscope camera was calibrated prior to the experiment by using the calibration method of Zhang [152].

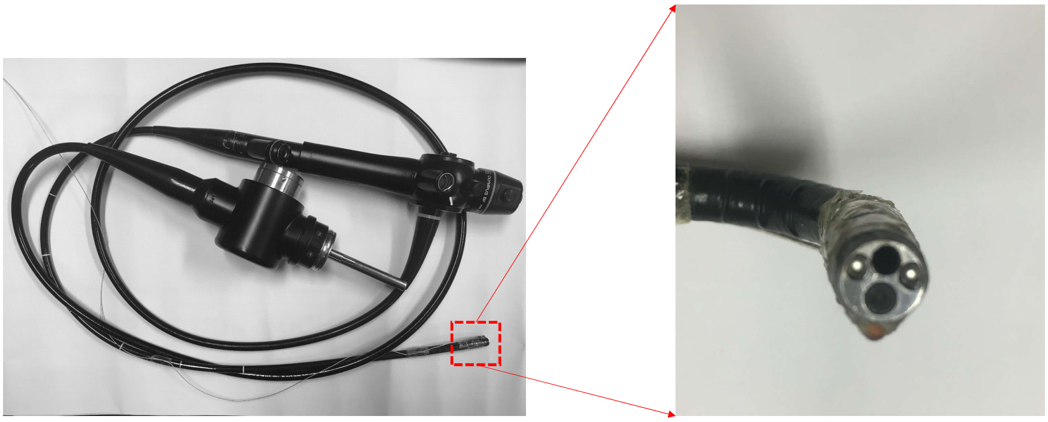
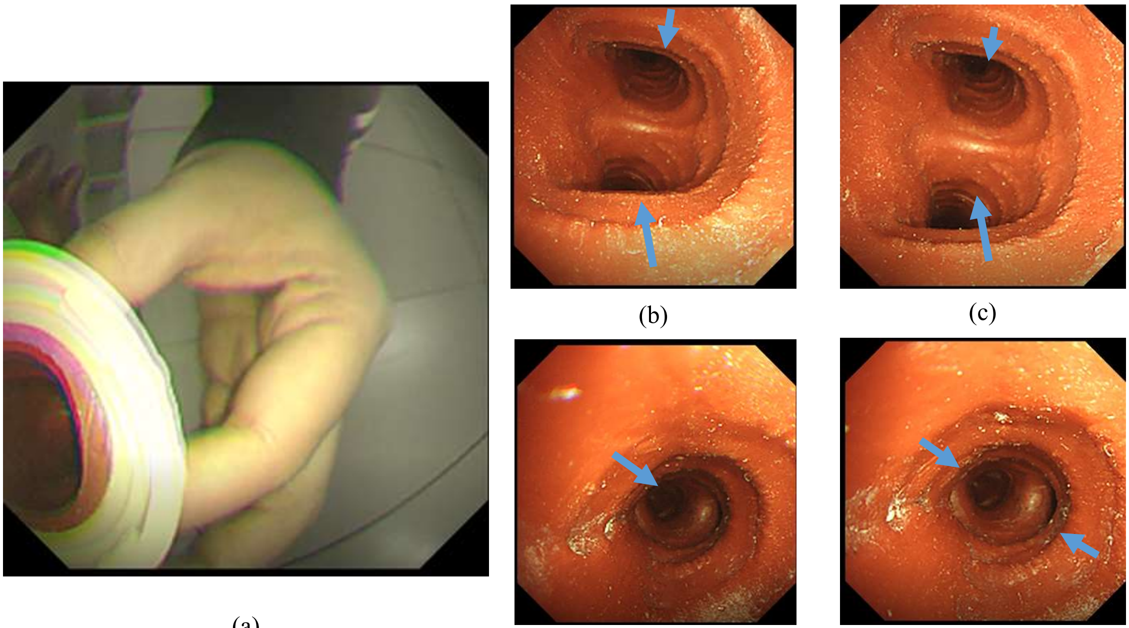


Figure 2.5: The bronchoscope used in our experiment

Table 2.4: Comparison of tracking results between original visual SLAM and proposed method (with ground truth)

Trial	ORB-SLAM [130]		Proposed method	
	TF	RMSE	TF	RMSE
1	111	-	3306	2.78
2	5106	4.13	5129	4.03
3	2934	4.08	2938	2.87
4	2078	2.61	2083	2.39
Average	2557	3.61	3364	3.02

TF: successfully tracked frames.



(a) (b) (c) (d) (e)

Figure 2.6: Illustration of the simulation of deformation. (a) observing from outside: adding additional force to the bronchus phantom using hand; (b)-(e) show the deformed bronchus from inside. The blue arrow shows the direction of the force approximately.

Table 2.5: Comparison of tracked frame number between original visual SLAM and proposed method (without ground truth)

Trial	Case	Frame No.	Moving trajectory	ORB-SLAM [130]	Proposed method
1	Case 1	3300	TR→ LMB	1903	4842
2	Case 1	5010	TR→ RMB	2305	3125
3	Case 2	7000	TR→ LMB	1023	1153
4	Case 2	7100	TR→ RMB→ TR→ LMB	1111	6982

TR: trachea; LMB: left main bronchus; RMB: right main bronchus; TF: successfully tracked frames.

2.7 Experiment results

2.7.1 Calibration results

The calibration procedure was performed prior to the validation experiment. This procedure was achieved by using the bronchoscope to capture the images of a chessboard. The chessboard was shown in Fig. 2.7. The calibration procedure was performed several times to obtain a high calibration accuracy. The rectified chessboard and RB image was shown in Figs. 2.7 and 2.8. As shown in Fig. 2.7, the lines in chessboard were parallax to each other after calibration, which meant the calibration procedure had good result.

2.7.2 Tracking results

To validate the tracking performance of the proposed method, the original and the improved visual SLAM were used to process the recorded videos. For better understand of the tracking procedure, the tracking procedure of the two methods were compared at the same frame, which were shown in Fig. 2.9 (left used the original visual SLAM and right used the proposed method). In this figure, the bronchoscopic images used for tracking, the 3D points used for pose estimation, the current position of the camera, etc. were shown. The current position of the camera is shown as a blue triangle. In this figure, the camera were located in the right main bronchus at the selected moment. while the camera pose from the original visual SLAM had gone out of the bronchus lumen.

The performance of visual SLAM in bronchoscope tracking were evaluated in three aspects: (1) the number of successfully tracked frames, (2) the tracking accuracy as compared to the ground truth, and (3) the processing time of each frame.

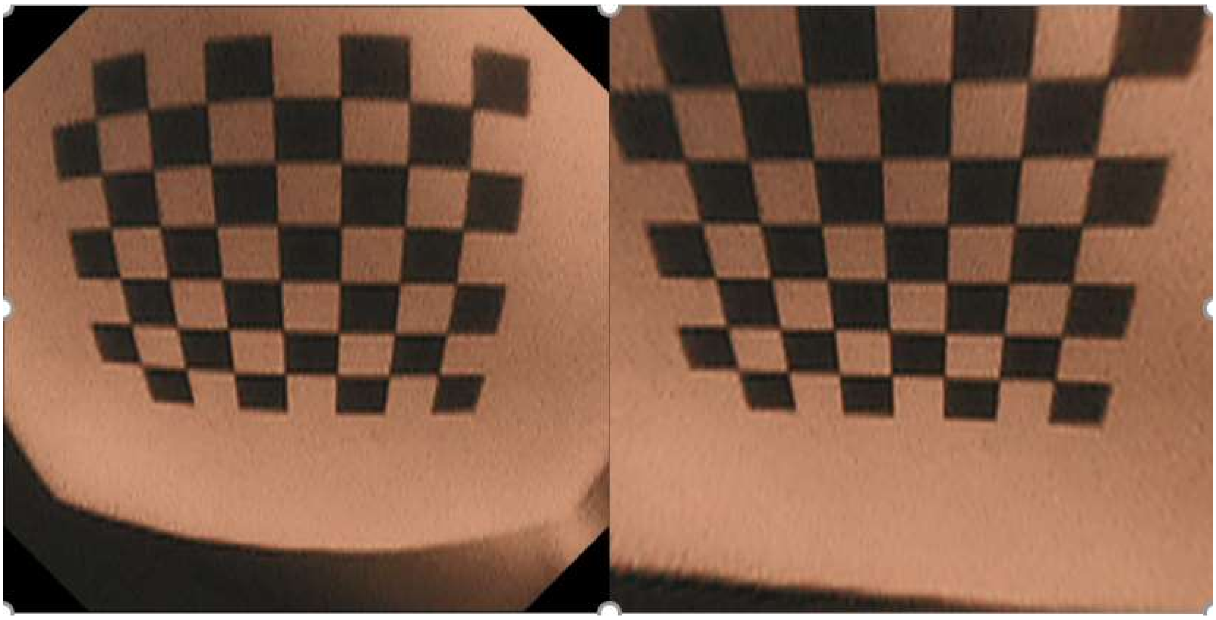


Figure 2.7: Chessboard used in calibration procedure and the rectified chessboard image.

Comparison of successfully tracked frames

The number of continuously tracked frames were recorded while processing each trail. More successfully tracked frames reflects a more robust visual SLAM as well as fewer tracking errors. Successfully tracked frames are defined as instances if the number of 3D-2D matches after pose estimation is larger than 30. The number of successfully tracked frames were shown in both the original SLAM and the improved visual SLAM in Table 2.4 and 2.5. The total number of successfully tracked frames was about 29,559 frames in all 37,863 frames, which corresponds the 78.1% of all acquired frames (including the trials in Table 2.4 and 2.5). The proposed method showed more successfully

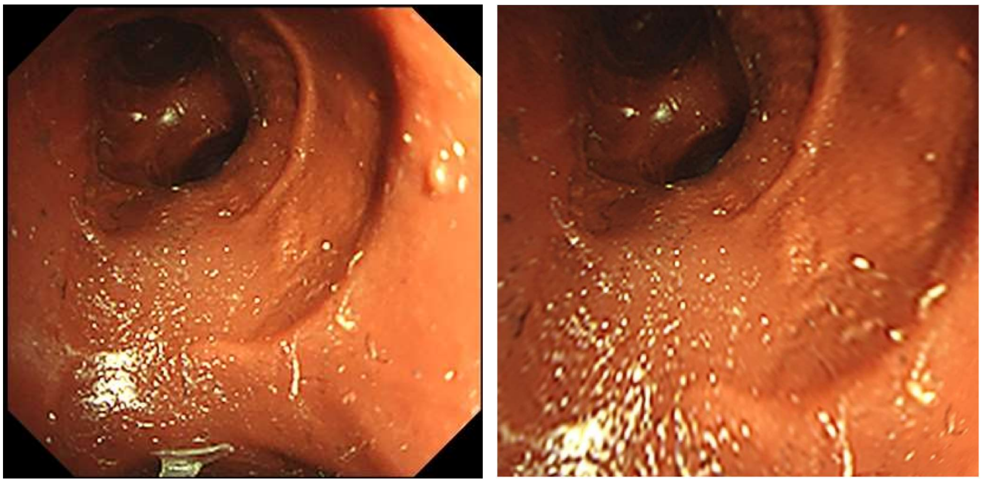


Figure 2.8: Example of the original RB image and the rectified RB image using calibration result.

tracked frames than the original method.

The number of successfully tracked frames in the other four trials were showed in Table 2.5. For these four trials, no ground truth of camera pose were created. Table 2.5 showed for the comparison of successfully tracked frames of the two methods.

Tracking accuracy

The accuracy of visual SLAM-based bronchoscope tracking was evaluated using the evaluation method described in Section 2.4.3. Root mean square error (RMSE) value was calculated with respect to the ground truth to evaluate the proposed method. These

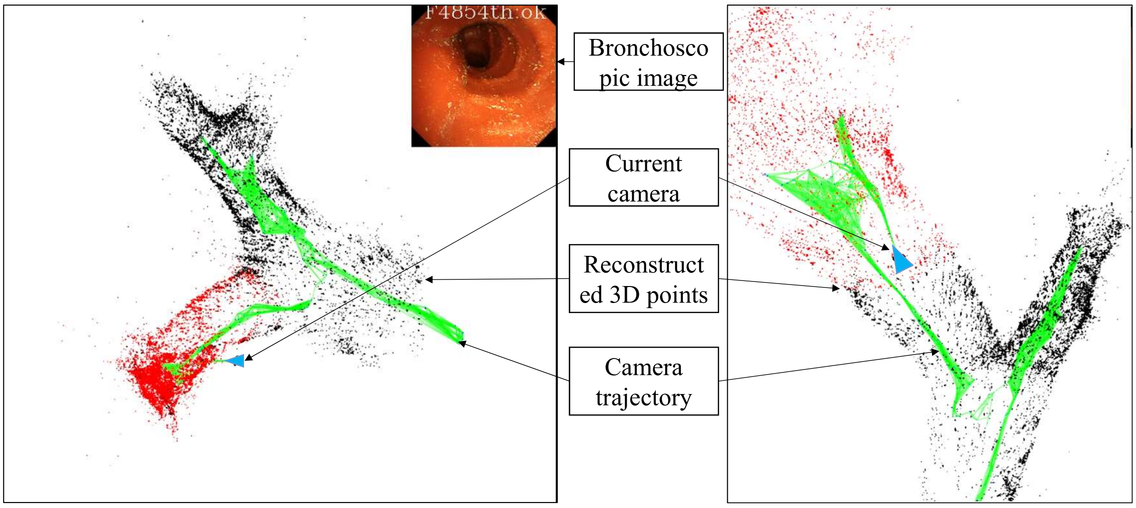


Figure 2.9: Comparison of tracking and reconstruction during visual SLAM-based bronchoscope tracking. The left image uses the original visual SLAM and the right image uses the proposed method. The reconstructed points are shown as black points; the camera trajectory is the green line; the current position of the camera is shown as a blue triangle; and the image in the middle top of the figure shows the k-th frame ($k = 4854$ in this trial). The camera pose from the original visual SLAM is obviously wrong because it had gone out of the bronchus.

results were shown in Table 2.4. The average RMSE of the proposed method was 3.02 mm, which outperformed the original ORB-SLAM (3.61 mm).

The transformed camera pose was visualized in the CT coordinates for quality validation Figs. 2.10 and 2.11, some of example images were shown in Fig. 2.12. Virtual bronchoscopic images were generated by using the volume rendering method [73, 74, 153]

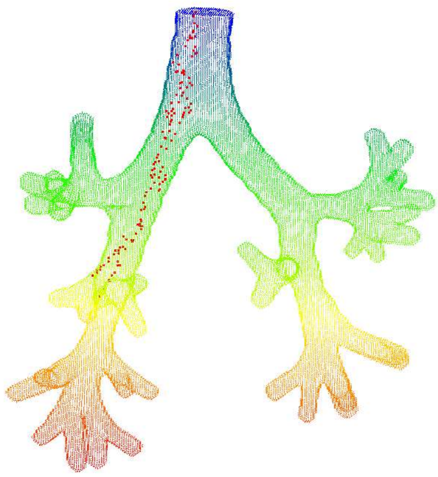


Figure 2.10: Case 1: estimated camera trajectory in bronchus (transformed to CT coordinate)

implemented in the software named “PLUTO”, as described by Nimura [151]. To visualize the bronchus surface, the color and opacity values for each voxel in the input CT image were set according to the CT value. The opacity value of voxels with values greater than -900 H.U. was set to nonzero and the opacity of the other voxels was set to zero. The color in these voxels was set to orange. Script is needed to write to load the transformed camera pose $\mathbf{P}_S^{(i)}$ and set the pose of the virtual bronchoscope in this software. The rendered VB image size was set to same as the size of the RB images (444×440 pixels). The field of view angle of the virtual bronchoscope was set to 100 degrees in diagonal. The virtual bronchoscopic images were generated and saved to files for validation.

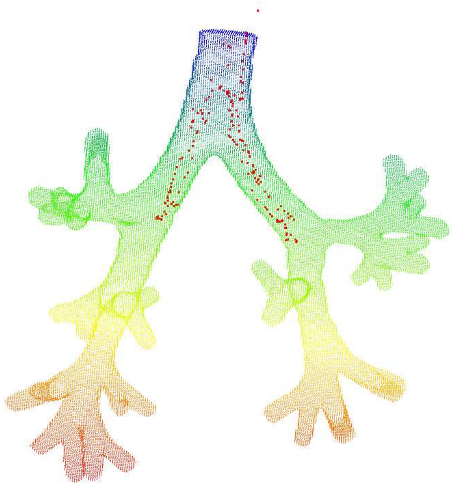


Figure 2.11: Case 4: estimated camera trajectory in bronchus (transformed to CT coordinate)

Processing time

The proposed method took about 80 ms to process one frame, which is a rate with great potential for real-time tracking. A possible way to decrease the processing time is to limit the number of used feature points in one frame to a reasonable number and implement the program onto a GPU board to accelerate the processing time. The proposed tracking scheme spent approximately 80 ms to estimate the camera pose of one frame, which is a little faster than the original visual SLAM, as shown in Fig. 2.13. This is because our proposed method found fewer 3D-2D matches during pose estimation due to a stricter criterion. Notice that the other settings of the two methods, such as the number of feature points expected to extract per frame, were set to the same values.

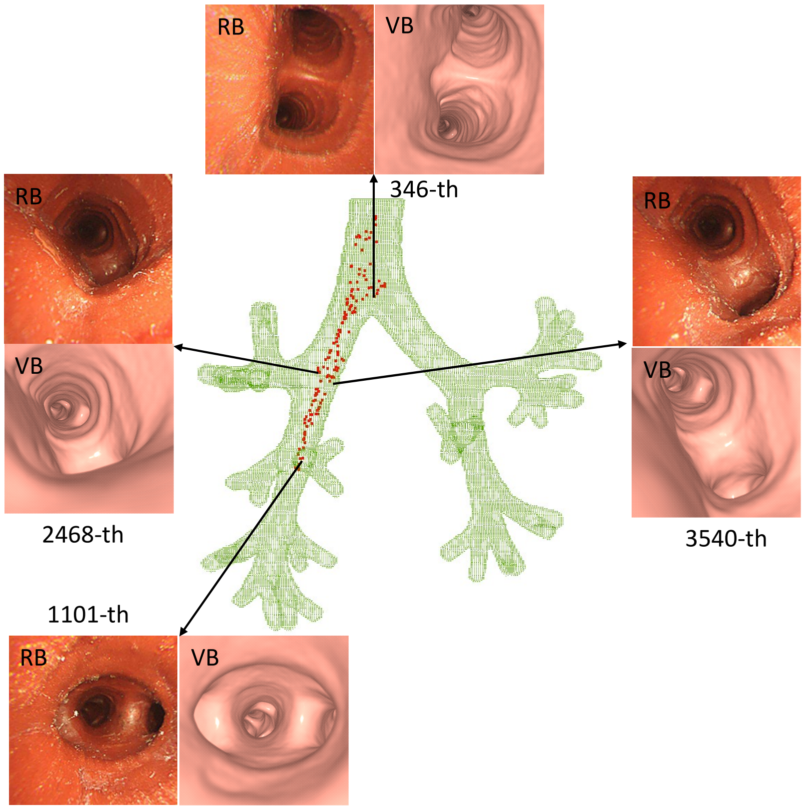


Figure 2.12: Camera trajectory from visual SLAM. The camera trajectory is shown in CT coordinates and the RB images together with VB images for comparison.

2.8 Discussions

Our proposed bronchoscope tracking scheme shows good tracking performance in phantom cases. The proposed method uses more accurate 3D-2D matches for tracking by limiting the size of a 3D region, thereby improving the accuracy of the tracking. Figure 2.14 showed an example of the comparison of two tracking procedures. More red area was observed in the left figure. The green line showed the camera trajectory, and the red points were the 3D points used in equation 2.1 in pose estimation. More red points meant that more 3D-2D matches were found and used. Based on the experimen-

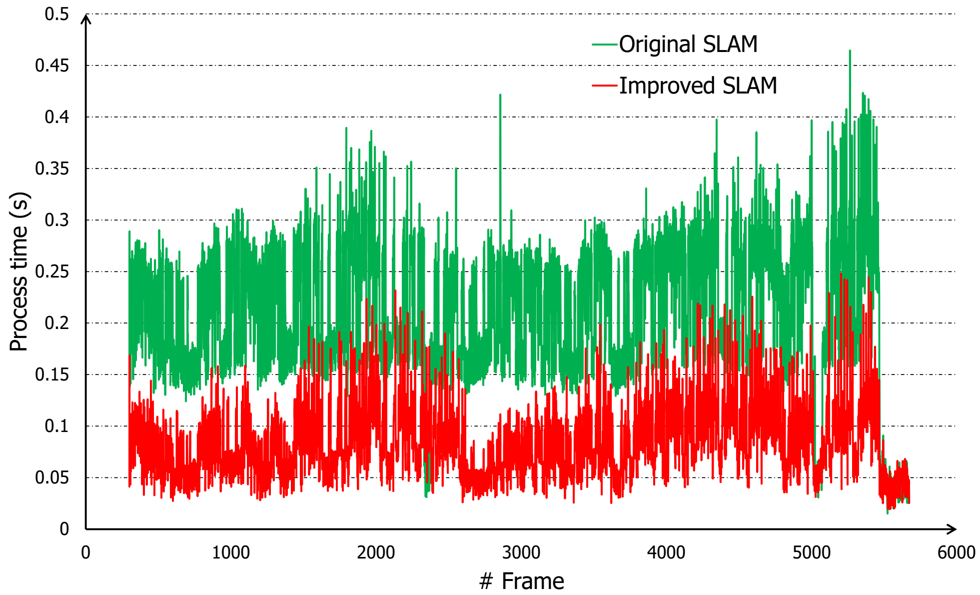


Figure 2.13: Processing time of each frame in original visual SLAM and proposed method for second trial. The average processing time of each frame was 0.21 s for the original SLAM and 0.08 s for the improved SLAM.

tal results, this tracking scheme can deal with a bronchus scene that lacks characteristic bronchus information (such as bifurcations), and it can deal with non-rigid deformation well. The tracking results in static and dynamic bronchus scenes demonstrated that the proposed method is suitable for bronchoscope tracking.

Tracking in scenes with deformation

The third and fourth trials in Table 2.4 showed that our proposed method behaves well in scenes with deformation. Both the third and fourth trials showed the RMSE values of the proposed method were smaller than those of the original visual SLAM. The

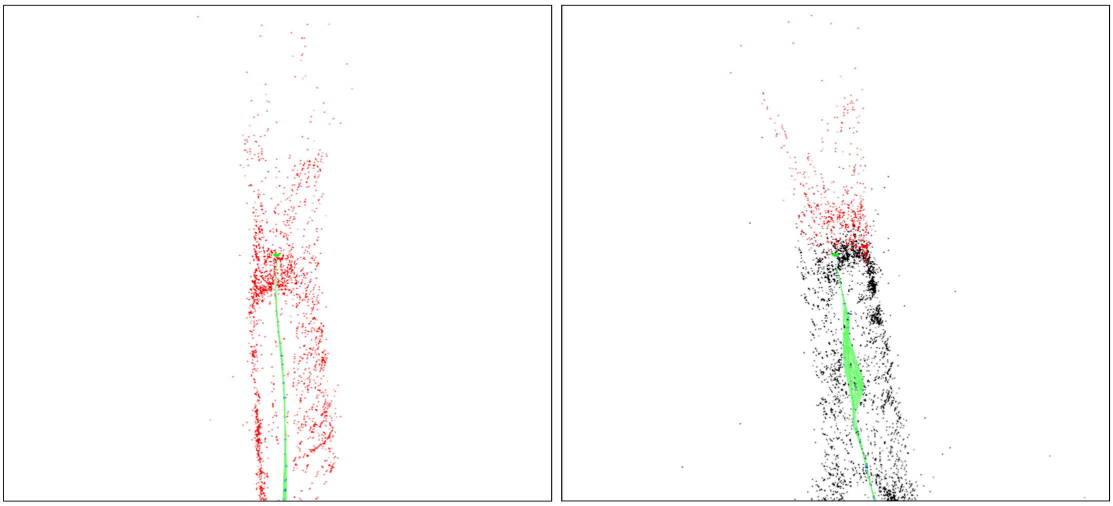


Figure 2.14: Screenshot of two methods while tracking (left, original visual SLAM; right improved visual SLAM). The black points in the two images show the reconstructed points; the green line shows the camera trajectory; and the red points show the 3D points used during pose estimation. As can be seen, there are more candidate 3D points for pose estimation in the left image than in the right image. Fewer red points means fewer 3D points were selected. As described in Section 2.4.2, an additional candidate point selection procedure is used to select the candidate 3D points. Since stricter conditions were used while selecting candidate points, the proposed method selected fewer points.

third trial had a larger RMSE value because the simulation of breath was larger than in the fourth trial.

However, the existing deformation lead to fewer reconstructed 3D points, which made the registration procedure more difficult than in the other trials, as it became

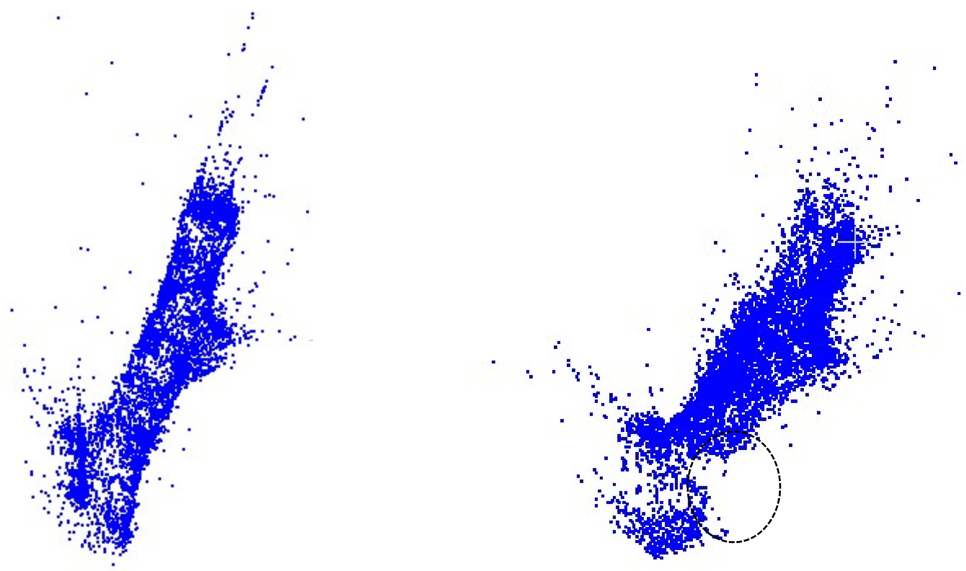


Figure 2.15: Comparison of reconstructed 3D points in first trial (left) and third trial (right). Due to the non-rigid deformation, some 3D points located in these region couldn't be reconstructed. Here, points located in black circle are lacking.

more difficult to select control points for registration. Figure 2.15 shows that there were fewer reconstructed 3D points in the fourth trial than in the first trial. In experiments, two bronchus phantoms for bronchoscope examination training were used to validate the proposed method. These phantoms do not simulate any anatomical structure changes, such as deformation by breathing motion or tumors. In future work, it should be considered to develop more precise bronchus phantoms that can simulate such deformation or anatomical changes, and use such phantoms for further experiments.

Investigation of tracking failure

The reason of tracking failure are investigated. As shown in Table 2.4, the RMSE of the second trial was larger than that of other trials. The image sequence of this trial was much longer than that of the other trials. The longer image sequence led to more outliers during tracking, and thus the larger accumulated error. It has been concluded that the outlier matches exist mainly in two procedures: 3D-2D matches in the pose estimation procedure and 2D-2D matches when creating more 3D points. For the generation of 3D-2D matches, more narrow criteria is used while searching for 3D-2D matches in coarse-to-fine pose estimation, and this lead to a small reprojection error while tracking as well as lower processing time (as shown in Fig. 2.13). The results of the proposed method demonstrate this.

Outliers existing in 2D-2D matches will lead to the outliers in reconstructed 3D points. As shown in Fig. 2.16, some points around the bronchi (as shown in the red circle) were reconstructed incorrectly. The 3D points are reconstructed by using 2D-2D matches in frames (see Section 2.4.1). The outliers in the 3D points may be caused by the mismatches in the 2D-2D matches. For the outliers in the red circle in Fig. 2.16, the 2D-2D match results in the frames that observed the bronchi area are investigated. An example of two frames used for reconstruction is shown in Fig. 2.17. This figure showed several matched feature points that were used for 3D reconstruction. The reconstructed 3D points had been checked against the narrow criteria before being accepted as new 3D points. However, the mismatches still could be found in the 2D-2D matches (as shown in the enlarged part of Fig. 2.17: in the 431^{st} frame, the two feature points are mostly in the same position; however, in the 400^{th} frame, they are apart from each other by several pixels). Therefore, an additional outlier elimination procedure may help to avoid such 3D points. If there are fewer outliers in the 3D points, according to equation 2.1, both the 3D position b_j and the camera pose will be more accurate. As a result, there will be an additional reduction of the RMSE value.

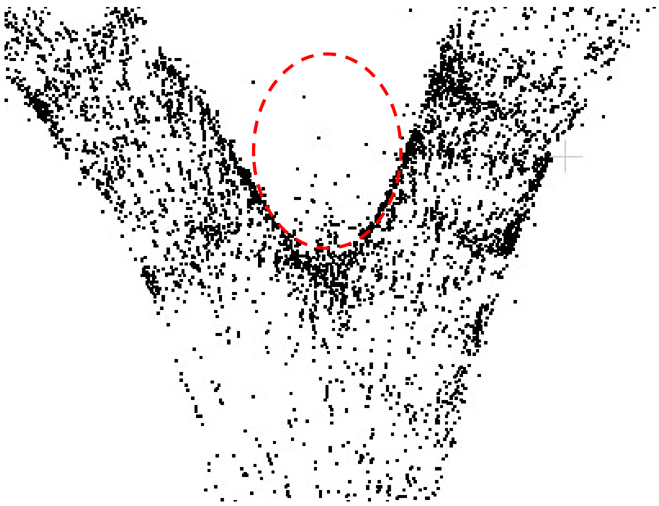


Figure 2.16: Example of outliers in point cloud of bronchus from visual SLAM of proposed method. The points in the red dotted line are considered as outliers.

2.8.1 Limitations of the proposed method

There are two limitations of the proposed method. One limitation is that the tracking accuracy will be decreased by large deformations. Visual SLAM-based tracking and reconstruction assumes that movement of the bronchus between two frames is very small. Therefore, deformation may cause errors in 3D reconstruction results. As a solution, compensation to deformation may be considered to further reduce its influence on bronchoscope tracking. Another limitation is that large specular reflection will also decrease the tracking accuracy when the bronchoscope hits or almost hits the bronchus wall; the strong light will lead to large specular reflection in the scene. It will become

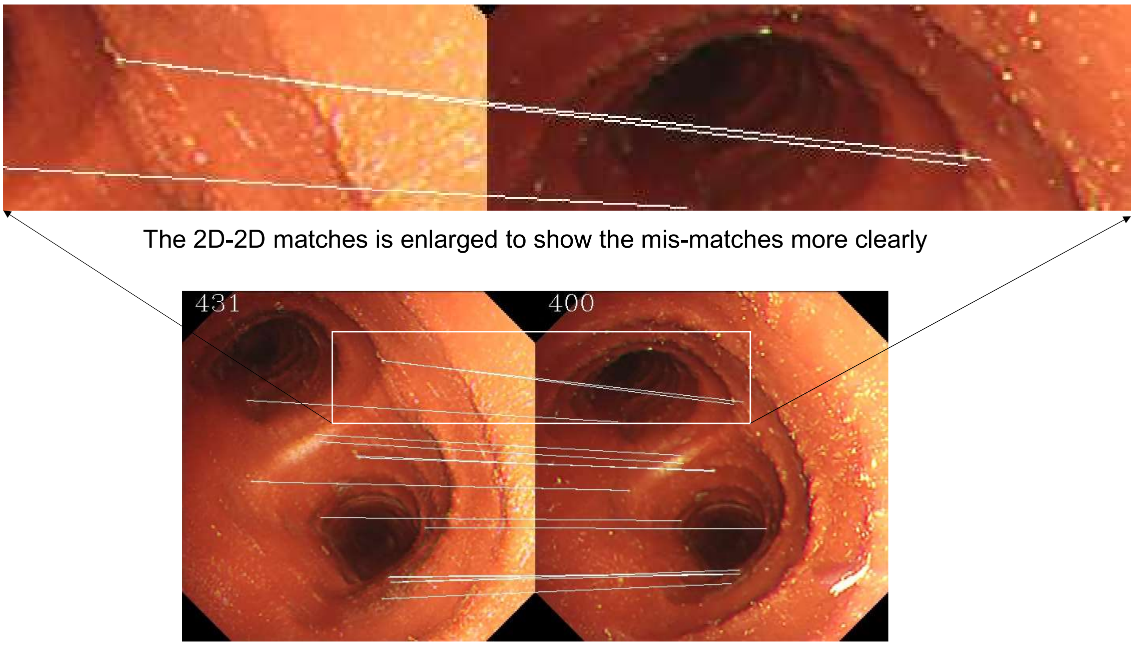


Figure 2.17: Two frames with 2D-2D matches drawing on them. The 2D-2D matches between the 400^{th} frame and the 431^{st} frame are picked as an example. A region of interest was enlarged for further examination. The two feature points in the 431^{st} frame were in same position, however the feature points in the 400^{th} frame were apart from each other. To improve the tracking accuracy, an additional outlier elimination procedure should be used to detect and remove the outlier matches.

difficult to extract feature points from these scenes. An additional specular detection procedure is needed to prevent these frames from being processed.

2.8.2 Future work

It is our next goal to validate the proposed method with *in-vivo* cases. Due to the lack of a camera calibration procedure as well as the lack of a ground truth in a clinical scene, the evaluation of the method *in-vivo* is still being developed. As mentioned in the discussion, the existing outliers in the 3D reconstructed points lead to a large tracking error. Therefore, the elimination of the existing outliers should also be considered in the future.

2.9 Summary

This chapter describes a new patient-specific bronchoscopy navigation scheme with bronchoscope tracking based on the output of improved visual simultaneous localization and mapping (SLAM). Visual SLAM uses only pure bronchoscopic images for tracking. Information from past images is stored to estimate the camera pose instead of using the image similarity between RB and VB, which will not be influenced by the difference between pre- and intra- operation information. Also, visual SLAM has better tracking results in areas with less structure. We improved the coarse-to-fine procedure in visual SLAM to find more accurate 3D–2D matches to make the technique more suitable for homogeneous bronchial lumen and evaluated the proposed method with more cases and trials; the tracking accuracy was evaluated by manually creating the ground truth of the camera pose. The transformed camera pose was visualized as virtual bronchoscopy for qualitative evaluation. The proposed method provides a new way for bronchoscope tracking and may be used in navigation in the future.

Chapter 3

Depth image-based bronchial orifice segmentation

This section describes the authors work on anatomical structure extraction in bronchus. In this section, a bronchial orifice (BO) segmentation method on RB video frames by using depth images is proposed. The BO structure is one of the anatomical characteristics in the bronchus, which is critical in clinical applications such as bronchus scene description and navigation path generation. This method utilizes the distance between the bronchoscope camera and the bronchus lumen, which is represented by a depth image. First several related works on image segmentation are introduced; next the depth image-based BO segmentation is introduced; and finally the experiment results are showed.

3.1 Background

During bronchoscopy, since the anatomical structure of the bronchus is complex and the bronchus scene is homogeneous, it is difficult to understand the current operative field.

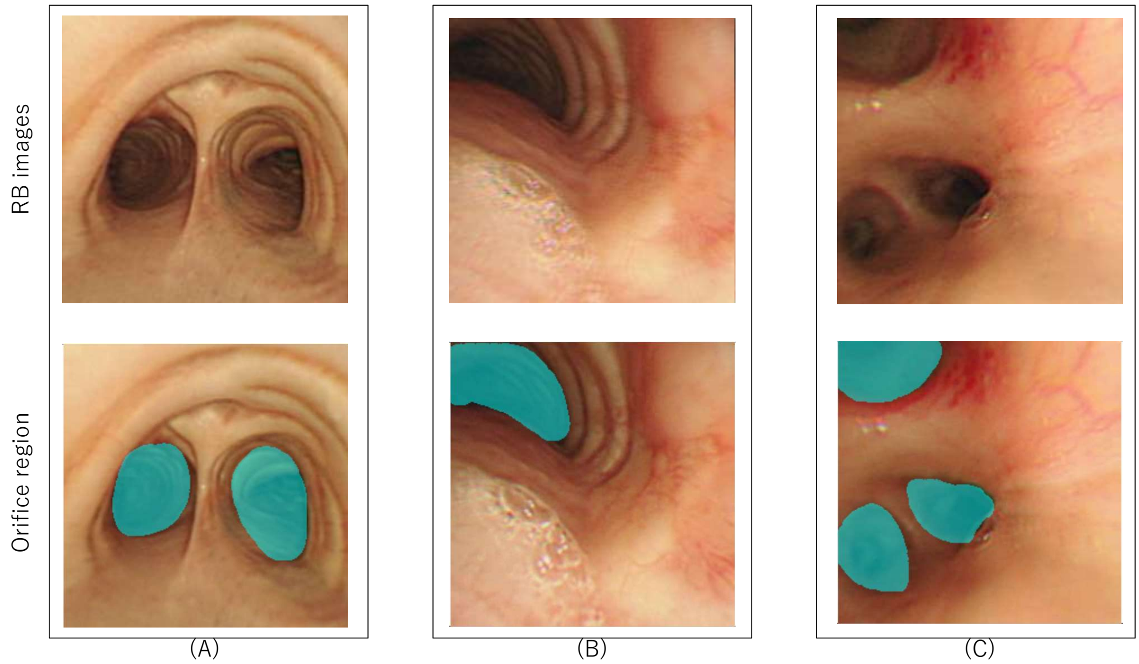


Figure 3.1: RB images obtained from different bronchus branches. The BO region are segmented manually. Different orifice shape and number is observed in different branches.

Therefore, scene understanding is an important task in surgeries. This research field contains many sub-tasks such as image classification, textual annotation, and object segmentation [112]. Due to the complexity of the bronchus scenes, object segmentation in the bronchus scenes, such as the segmentation of the anatomical structure (the BO and the carina) and surgical tools is more meaningful. Moreover, precise BO segmentation will benefit the development of clinical applications such as the generation of navigation path, quality assessment of a surgery, etc. [154, 155]. An example of the orifice region is shown in Fig. 3.1. The expected orifice region are manually segmented.

3.2 Previous work

Several research groups work on scene understanding by applying the segmentation method on endoscopic videos. These methods are roughly classified into classical segmentation method and deep learning-based method, which are introduced in the following section.

3.2.1 Classical segmentation method

Classical method-based use the appearance of endoscopic images and the geometry shape of the surgery tools for segmentation. Tonet et al. segmented the laparoscopic instrument from the endoscopic image by considering the geometric features and the color strip of the tool [156], the color strip of the instrument is obtained prior to the segmentation task. Zabulis et al. proposed an improved version of the mean shift-based method to detect the orifice region in the image obtained from a capsule endoscope [157]. Their method extracted up to two orifices in an endoscopic image. Sanchez et al. showed an image appearance-based method to find the center of the lumen [154]. They used the intensity of pixels and its gradient in an endoscopic image to construct a 2D feature space. They observed the distribution of the features and used the found local maxima in each cluster of feature space as the lumen center. The parameters used in this procedure were concluded from a slight dataset. However, lumen shapes may change due to the deformation of the organ, which may decrease the performance of the algorithm. Asari proposed a gastrointestinal lumen segmentation method by using the differential region growing method (DRG) on the gray-scale image [158]. A two-stage DRG is used to segment the lumen and shows good performance.

3.2.2 Deep learning for segmentation

The deep learning-based segmentation uses the manually created data to train a model to represent the feature of the desired region. Morimitsu et al. used a deep learning model named LSTM U-Net to segment the blood vessels in laparoscopic videos. This network contiguously estimated the inferior mesenteric artery (IMA) region by using the previous frames [159]. Pakhomov et al. used an improved fully convolutional network (FCN) to segment the surgery tool's shaft, tool's manipulator, and background from real laparoscopic video [160]. Bravo et al. used two convolutional neural networks (CNN) to detect the polyp region in the colonoscopic video [161]. The segmentation procedure was divided into polyp classification and localization parts. Images showing high probabilities of having polyp were used as the input of the localization part for a robust segmentation. However, conventional image appearance-based methods have many shortcomings. Firstly, the color image-based methods are sensitive to the changes in illumination. Secondly, they may have a poor performance in complex scenes such as bubble exists or the deformation of the organ [108, 155]. And deep learning-based methods need additional labor in the preparation of the training data.

3.3 Purpose of this Chapter

In this Chapter, a new BO segmentation method is proposed based on the depth image estimated from deep learning. The depth image is estimated from an image-to-image network named cycle generative adversarial network (CycleGAN) [162] from a bronchoscopic video. Then the depth image pixels are used to decide the threshold value for image binarization since the depth range differs in branches. The obtained binary image shows the BO region.

3.4 Proposed method

The goal of this topic is to segment the BO regions on RB video frames. With an assumption that the bronchoscope will always observe the orifice region during the examination, the distance between the bronchoscope and the inner surface of the lumen, which is represented by depth images, is used to extract the orifice region in the RB image. As shown in Fig. 3.2, the proposed method mainly consists of two steps: (1) CycleGAN-based depth image estimation, which estimates depth images from RB images, and (2) depth image-based orifice region segmentation, which finds the orifice region by binarizing the depth image based on projection-based threshold decision.

3.4.1 Depth image estimation

Due to the fact that the depth image is unavailable from the bronchoscope, the depth images are estimated from the RB images. An image-to-image translation network named cycle generative adversarial network (CycleGAN) [162] is utilized to estimate depth images. This network has shown good performance in previous work for the generation of real colonoscopic images from the virtual colonoscopic images [163]. Furthermore, it doesn't need the paired data for training, which suits the bronchus scene. The procedure of the depth image estimation is introduced from three aspects: (1) introduction of CycleGAN for image domain translation; (2) preparation of training data, and (3) CycleGAN-based depth image estimation.

Introduction of CycleGAN for image domain translation

CycleGAN consists of several networks for image domain translation. It contains two image generators and two image discriminators. Image generators are based on an improved U-Net[164] (each convolution layer is followed by the instance normalisation

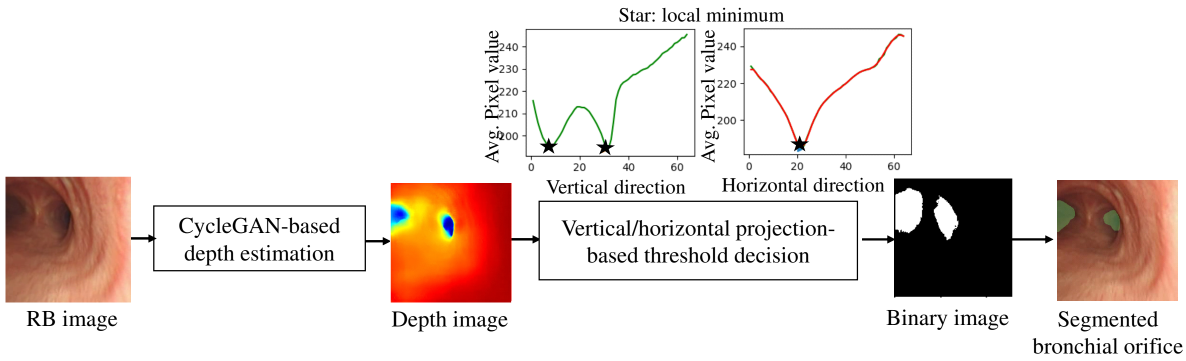


Figure 3.2: Flowchart of the proposed method for BO segmentation. The RB image is used as the input, depth image is estimated using CycleGAN; the orifice region is segmented using depth image; the white region show the segmented orifice.

[165]); the discriminators are based on a 4-layer convolutional neural network (CNN). The C-to-D image generator manages the translation from color RB images to depth images and the D-to-C image generator manages the translation from depth images to color RB images. Two discriminators are used to judge whether an image is generated image (fake image) or real image (prepared data) in two image domain, respectively (Fig. 3.3 (a)). CycleGAN contains two sub procedures: the forward procedure and backward procedure. In forward procedure, an RB image in RB image domain $\mathbb{R}$ is translated into a fake depth image in depth image domain $\mathbb{D}$, the fake depth image is translated back to $\mathbb{R}$ as a reconstructed RB image; in backward procedure, a depth image in $\mathbb{D}$ is translated into $\mathbb{R}$ as a fake RB image, this fake image is translated back to $\mathbb{D}$ as a reconstructed depth image. This two procedure are shown in Fig. 3.3 (b). Cycle-consistency loss is used to encourage the appearance of the reconstructed image similar to the original image. The training procedure of CycleGAN is to find a overall loss that contains the aforementioned loss functions (loss of two generators, loss of two discriminators and two cycle-consistency loss). A detailed description of CycleGAN is reported in literature [162, 163].

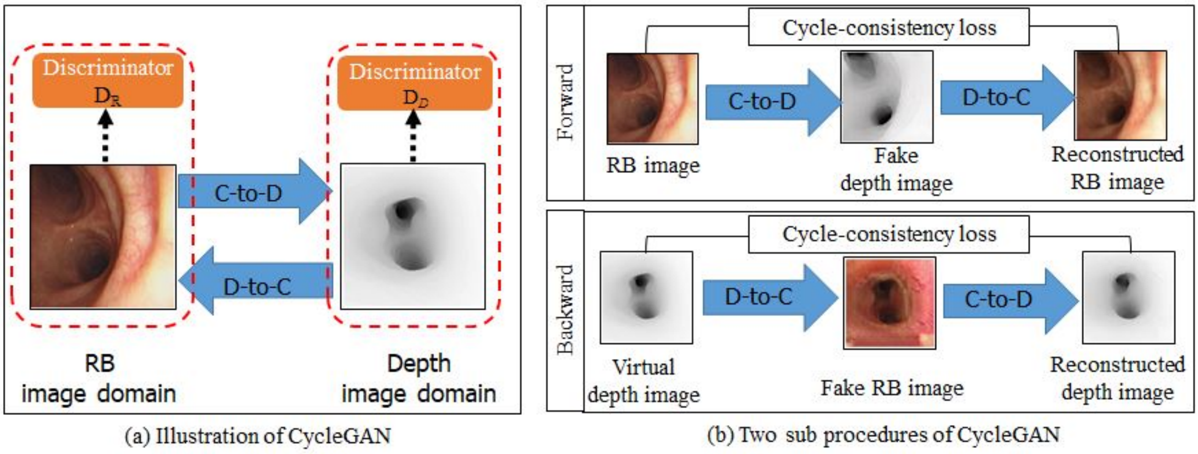


Figure 3.3: Illustration of CycleGAN and its two sub procedures. Generator is used to translate an image from original image domain to target domain. Discriminator is used to judge if an image comes from prepared data or generated by generator. Two sub procedure are the forward and backward procedure. Cycle-consistency loss is used to encourage the appearance of the reconstructed image similar to the original image.

Training dataset preparation

We defined two image domains, RB image domain $\mathcal{R}$ and depth image domain $\mathcal{D}$, for CycleGAN. The preparation of the training images in $\mathcal{R}$ and $\mathcal{D}$ is described in the following section.

Preparation of RB images

The RB images were obtained from RB videos. Since the original images from a bronchoscope have different imaging conditions, we adopt a data-cleansing procedure [163] to remove images with poor imaging quality. According to the literature [163], the data-cleansing procedure benefits the quality of image domain translation. An RB image is removed if one of the following events occurs: observation of a strange color; the camera hits (or nearly hits) the bronchial lumen (high specular reflections); the presence of bubbles. We show several examples of the removed and the remaining images in Fig. 3.4.

Preparation of depth images

The depth images are the virtual depth images generated by a virtual bronchoscopy system [74] from preoperative CT images. A virtual depth image is an image that stores the distances from the virtual bronchoscope's viewpoint to the surface of the bronchus lumen in CT images. The bronchoscope is moved along the centerline of the bronchus to generate virtual depth images at various viewpoints. We explore at least three branching levels when generating the virtual depth images: the trachea, the left and right main bronchi, and the truncus intermedius of the bronchus. The procedure of generating virtual depth image is shown in Fig. 3.5. An example of the depth image is shown in Fig. 3.4 (c).

CycleGAN for RB-to-depth image translation

We use CycleGAN to estimate the mapping X from the RB image domain $\mathcal{R}$ to the depth image domain $\mathcal{D}$. During CycleGAN's training procedure, an image $\mathbf{R}$ in $\mathcal{R}$ is translated to an image $\hat{\mathbf{D}}$ in $\mathcal{D}$ using a U-Net-like generator [163]. Since there are two types of depth images (translated by a generator and prepared in advance), an adversarial discriminator is used to distinguish which type of depth image belongs to. The generator is trained to change the images' appearance in an original domain (e.g., $\mathcal{R}$) to that of a target domain (e.g., $\mathcal{D}$). Details on CycleGAN for image translation can be found in the literature [162, 163].

CycleGAN-based depth image estimation

The prepared training data are used to train CycleGAN. The trained C-to-D image generator is used to estimate depth image from RB image.

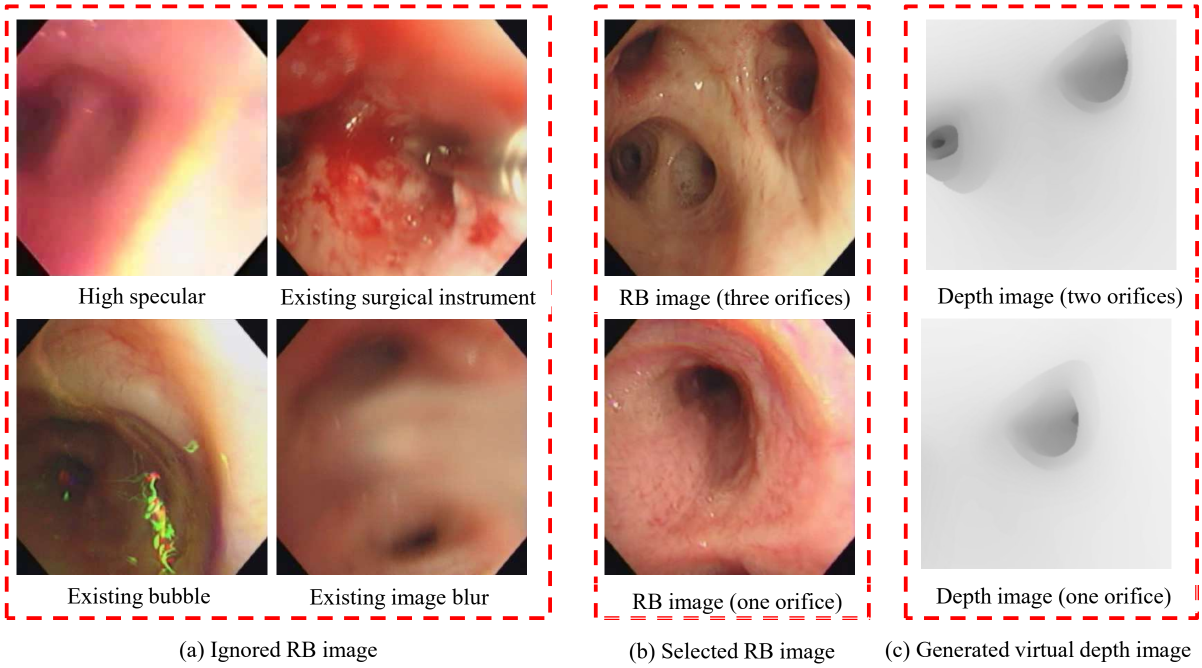


Figure 3.4: Several examples of ignored images and selected images while preparing training data. (a) ignored RB images; (b) selected RB images; (c) generated virtual depth images. The ignored images are the images existing poor image condition such as image blur, bubble and so on. The selected images are images showing good image condition. The depth images are generated from CT images.

3.4.2 Depth image-based BO segmentation

The depth image is used for the segmentation of the BOs. As a matter of fact, points located in the orifice region are farther from the bronchoscope than other regions. Therefore, a threshold value is used to process the depth image in order to find the orifice region. Furthermore, since the depth range differs a lot in different branching levels, an image pixel projection-based method is adopted to decide the threshold individually for each depth image instead of using a fixed value.

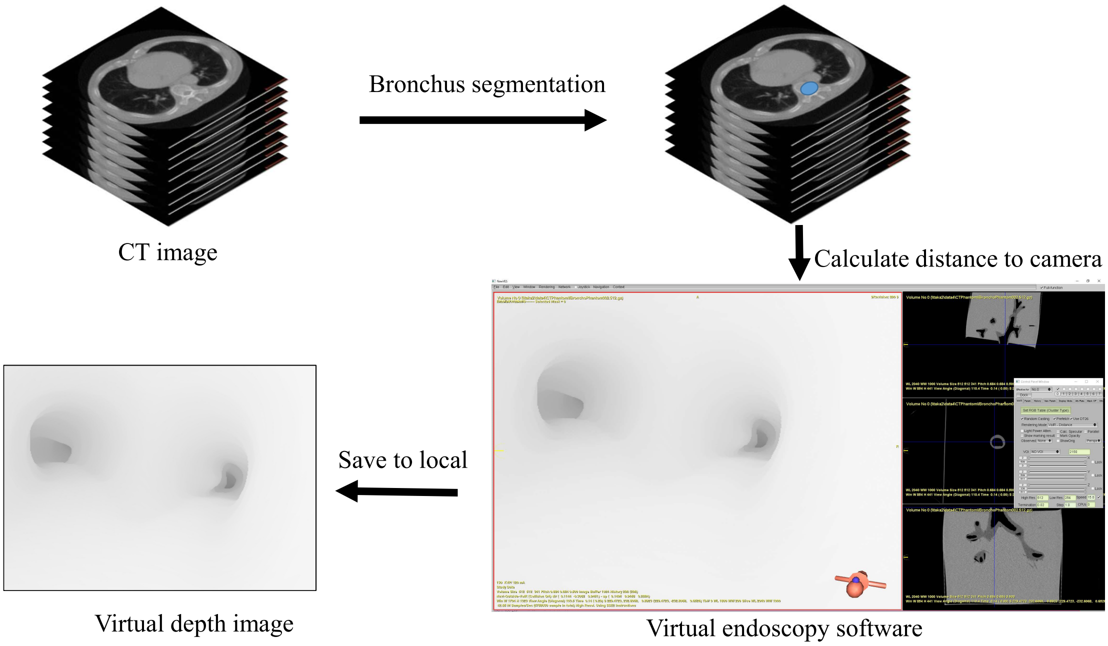


Figure 3.5: Illustration of the generation of virtual depth image from CT image. A virtual endoscopy software is used to check the bronchus region while the generation of virtual depth image.

There are three steps to decide the threshold for each depth image. The first step is to obtain the distribution of pixel values in the depth image at vertical and horizontal directions, which is achieved by calculating the average projection profile in two directions. The width of the depth image is denoted as W and the height is denoted as H. The average projection profile of pixel value in the horizontal direction is obtained by using $I_x(m) = \sum_{n=1}^{H} \mathbf{D}(m,n)$ and in the vertical direction is $I_y(n) = \sum_{m=1}^{W} \mathbf{D}(m,n)$, where $\mathbf{D}(m,n)$ is the pixel value at the m-th column and the n-th row in depth image $\mathbf{D}$. Two curves $I_x(m)$ and $I_y(n)$ show the distribution of the values. The second step is to find the local minimum values in each curve. Since the orifice region is farther from bronchoscope, the orifice region is observed as lower pixel value area in projection curve, which corresponds to the wave trough of the curves. The carina region corresponds to the peak between trough. An example is shown in Fig. 3.6. In this figure, two orifice regions are observed in the RB image, which correspond to two 'darker' regions in the depth image. The average projection pixel value in vertical and horizontal direction are calculated in depth image. The wave trough region in the projected profile correspond to the 'darker' region in depth image. The third step is to decide the threshold value to extract the orifice regions. The local minimum values in $I_x(m)$ and $I_y(n)$ are sought by comparing the neighborhood value on curves. The found value set is marked as $\mathbf{V} = \{V_1, V_2, \cdots, V_N\}$. The threshold value th is decided by using equation $th = \max(\mathbf{V})$. We use a smoothing procedure before find the local minima. Note that this procedure is not necessary. An example of curves after smoothing procedure is shown in Fig. 3.6 (c). The smooth procedure removes the jitters in curve and decrease the possibility of finding the wrong local minimum value which do not correspond to orifice regions. The found local minimum value are marked in stars on curves.

The found threshold value is used to perform image binarization on the depth image. The orifice regions are considered as the regions in which the pixel intensity is not equal to zero.

3.5 Experiments and results

The proposed method was implemented in the Ubuntu 16.04 system with an Intel Xeon Gold 6134 processor (8 cores at 3.2 GHz) and integrates with an NVIDIA Tesla V100 GPU card. We use three *in-vivo* cases and one *ex-vivo* case for the training. The total number of the RB images used in experiment was 4,398 (3,131 for training and 1,267 for testing) and the total number of the depth images was 4,151. We crop the obtained

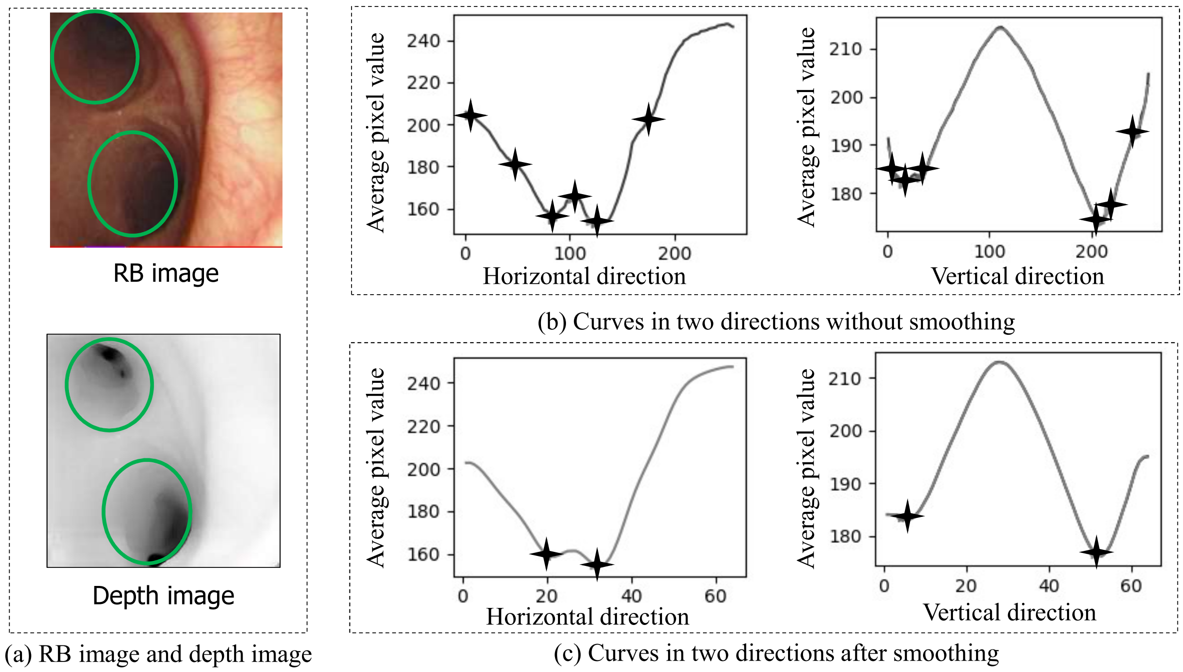


Figure 3.6: RB image, the corresponding depth image and the projection profile in two directions before/after smoothing. The found local minimum values are marked on curves. Curves after smoothing show accurate results.

RB images to 256×256 pixels for training. We set the training epoch to 200 with a 10 batch size.

3.5.1 Quantitative evaluation

We used the proposed method to perform the BO segmentation from *in-vivo* videos obtained during the examination. We manually labelled BO regions as ground truth for quantitative evaluation and implemented two previous methods to segment the BO region from RB images. These two methods are: method (1) use Otsu thresholds [166] on a grayscale RB image to obtain the threshold value for binarization; method (2) use U-Net [164] trained from color RB images and manually created label images. U-Net was implemented with the Dice loss as the loss function. We created a tiny dataset of three *in-vivo* cases, including about 130 frames, to train the U-Net model.

We compared the performance of these methods for BO segmentation. The Dice score was used to measure the accuracy of these methods. The Dice values in four cases of different methods was shown in Table. 3.1. The average Dice score of the proposed method was 77% in 219 frames.

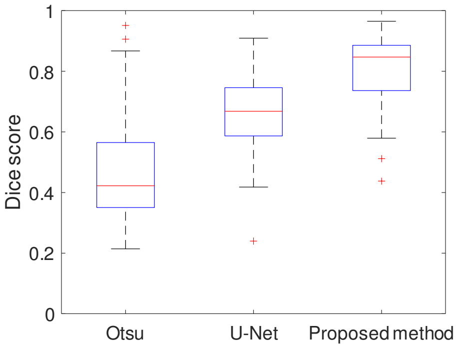
3.5.2 Qualitative evaluation

For qualitative evaluation, we selected several frames to investigate the segmentation results in various scenes. We show the segmented results of the first and the second cases in Fig. 3.8. The BO regions segmented by the proposed method were marked blue in RB images.

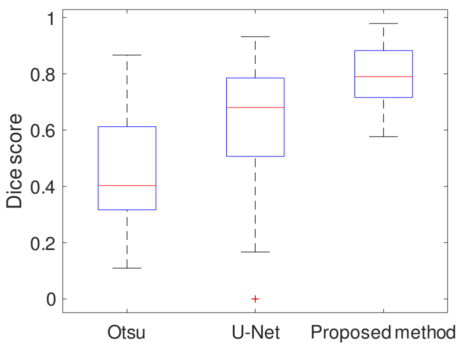
To show the segmentation results of different methods, we randomly selected several frames of the third and the fourth cases. The selected frames and the segmentation results of different methods were shown in Fig. 3.9. We showed the RB images, the results from the proposed method (marked as 'Ours'), the results from the Otsu (marked as 'Otsu'); the results from U-Net-based segmentation (marked as 'U-Net'); and ground

Table 3.1: Quantitative evaluation result of different methods

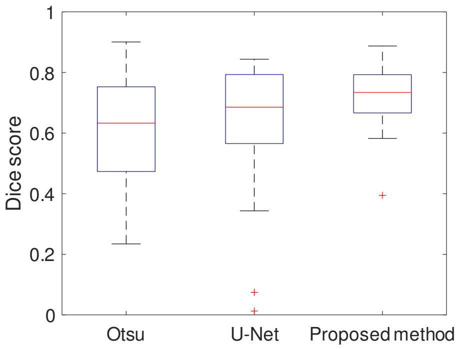
Case	Frame number	Otsu [166]	U-Net [164]	Proposed method
1	77	0.46	0.67	**0.81**
2	68	0.46	0.60	**0.79**
3	27	0.60	0.63	**0.72**
4	47	0.54	0.62	**0.76**
Avg.	54	0.52	0.63	**0.77**



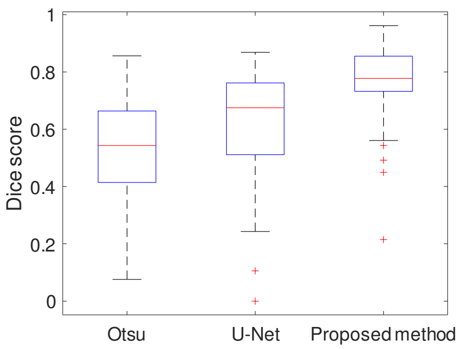
[1] Case 1



[2] Case 2



[3] Case 3



[4] Case 4

Figure 3.7: Boxplots of Dice score from different segmentation methods in four cases. The proposed method shows higher average value than other two methods.

truths (marked as 'GT').

3.5.3 Processing time of the proposed method

We measured the processing time of the proposed method in about 20000 frames. The processing time per frame contains the generation of the depth image, threshold decision, and visualization of the segmentation result. The average computation time for each part was 0.005 seconds (GPU), 0.002 seconds (CPU), and 0.006 seconds (CPU). The total processing time per frame was 0.013 seconds.

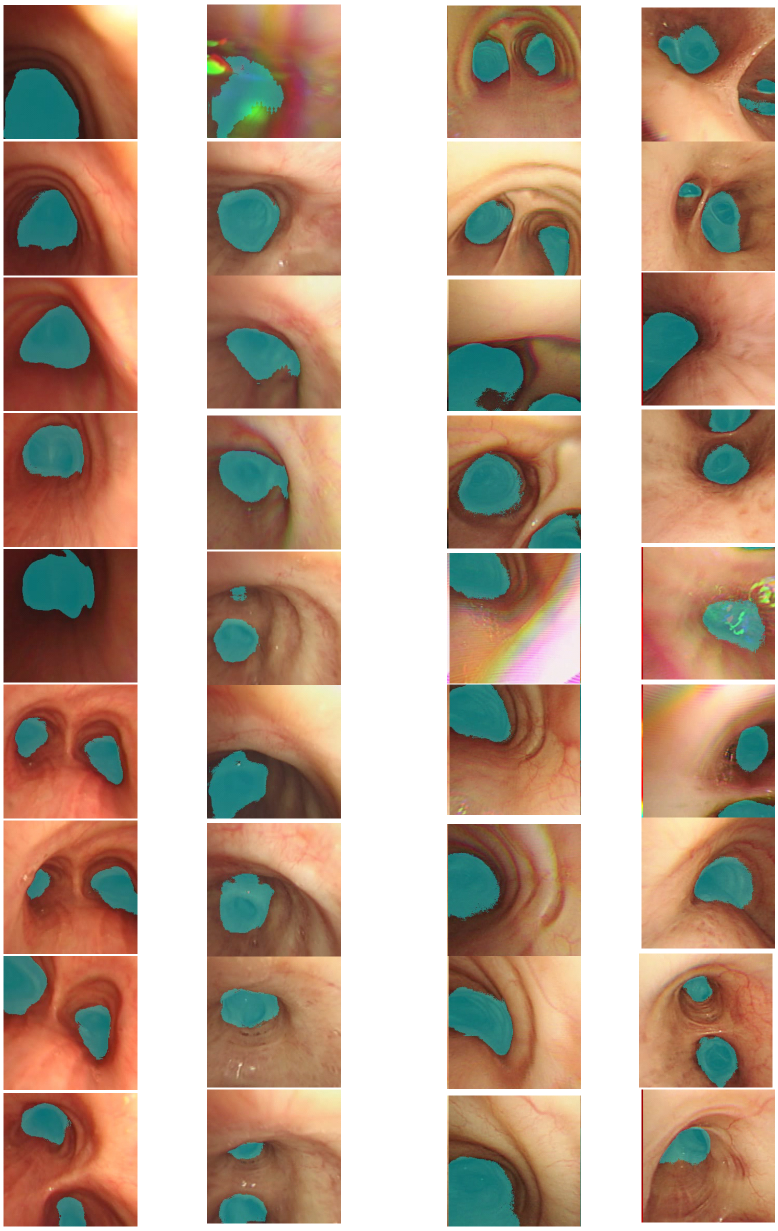


Figure 3.8: Examples of segmented BO regions in four cases. Each row indicates same *in-vivo* case. Numbers under each image shows frame number. The number of BOs ranges from one to three. Most of the BO region was well segmented by the proposed method.

3.6 Discussion

As shown in Table 3.1 and Fig. 3.7, the proposed method has the highest Dice score among the three methods. The proposed method shows the best performance of the three methods in our experiment. Meanwhile, as shown in Fig. 3.7, the proposed method shows smallest variance in three methods. We conclude with the following two advantages of the proposed method comparing to the previous methods.

3.6.1 Advantage of the proposed method

Accurate segmentation for each BO

The proposed method can segment each BO region individually. There are two or three BO regions in typical bronchoscopic scene. As shown in Figs. 3.8 and 3.9, the proposed method segmented almost each BO region while the previous methods did not separate each BO region. From Fig. 3.9, we can observe that the Otsu-based segmentation finds some false positive regions as the BO region. We think this is because Otsu method did not obtain appropriate threshold to segment the BO regions for bronchus scene. The BO regions from the U-Net-based method behaves better than the Otsu-based segmentation. However, U-Net-based segmentation cannot identify each BO region as shown in the frame (3-3690) and (4-6426) of Fig. 3.9. Moreover, it is necessary to prepare the training data, which costs huge labor to label the BO region manually. It may have better performance if more training data are used. The proposed method shows the best segmentation result. We think this is because the proposed method is based on the depth image instead of the color image. In the depth images, the BO regions tend to have lower pixel value than the carina region. Since the threshold used for segmentation is lower than the pixels in the carina region, only the region farther than the carina region is segmented as BO region.

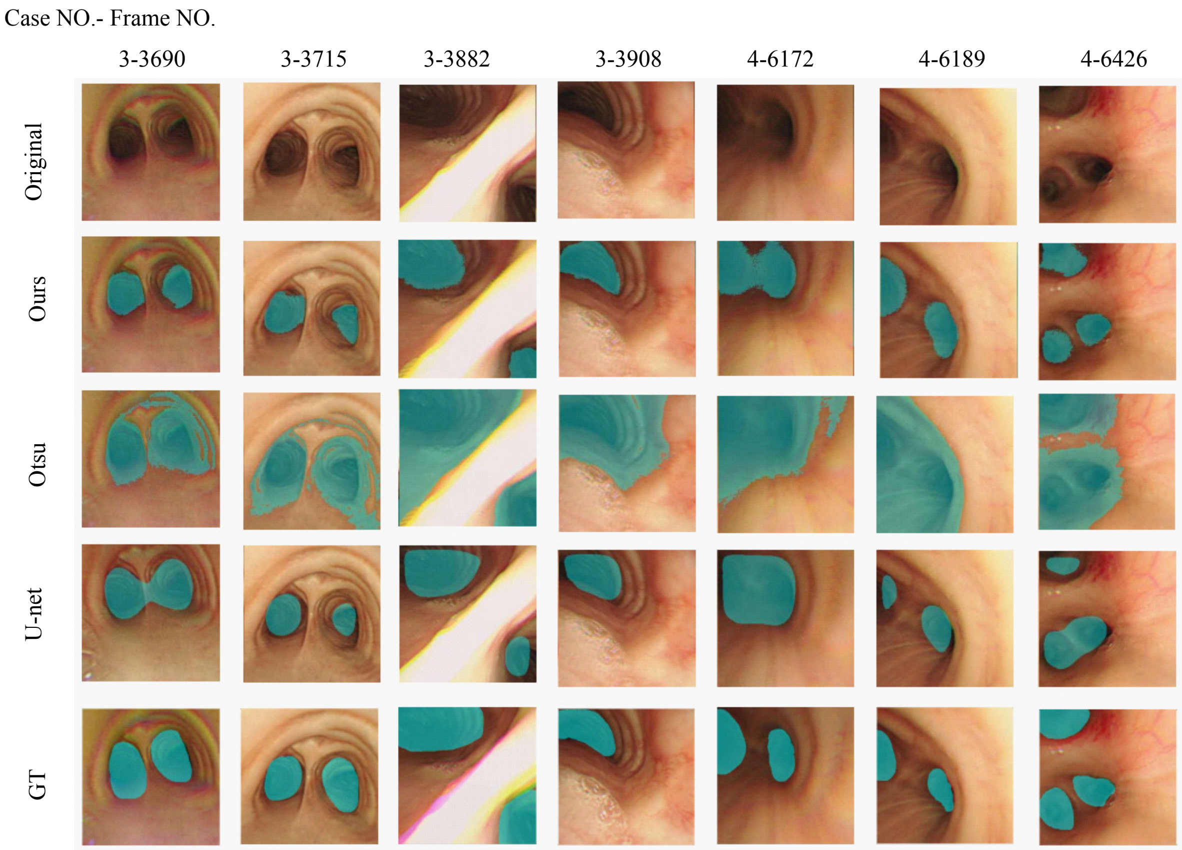


Figure 3.9: Comparison of the results from the different segmentation methods.Frames are selected from cases one and two. We showed the original image (Original), results from the proposed method (Ours), results from the Otsu method (Otsu), results from U-Net (U-Net) and ground truth (GT). The results of the proposed method have the most similar appearance with ground truth.

Robust to the complex image conditions

The proposed method shows good segmentation results in images with poor image quality. We think it is the benefit of using depth images. The influence of poor image quality has disappeared in the depth image. We show two example of RB images with poor image quality, the segmentation results, the generated depth images, and results from other methods in Fig. 3.10. The first row shows an image existing image blur

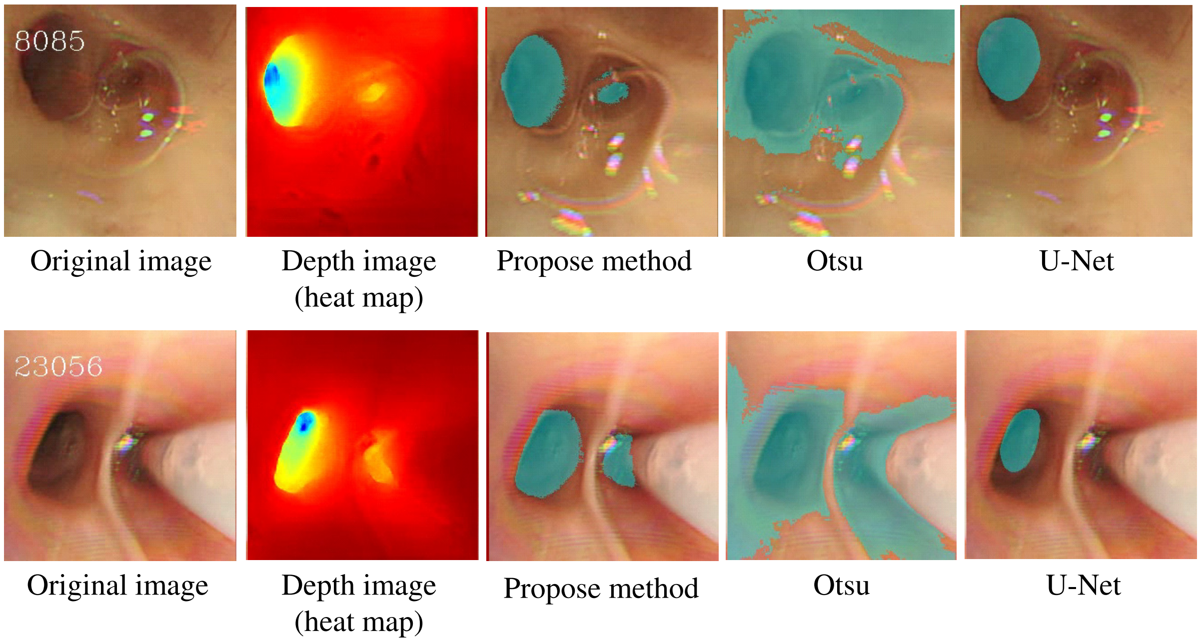


Figure 3.10: Examples of segmented BO regions in images showing poor image quality. Image in the first row contains bubble; image in the second row contains surgical tool. We show the original RB image, estimated depth image in heat map, result from the proposed method, result from Otsu method and result from U-Net. The proposed method show better segmentation result.

and strange color. The second row shows an image existing surgical tool. We also show the generated depth images of each image. These poor image conditions have disappeared in the depth images. Therefore, the poor image conditions will not influence the threshold procedure, the proposed method obtains good segmentation result.

3.6.2 Limitation of the proposed method

One of the shortcomings of the proposed method is that it doesn't judge whether the BO region exists or not. Since we assume the BO region always appears in the RB image, the proposed method will always try to find the BO region in the current view, as shown in Figs. 3.11 (a) and (b). A possible solution is to use an online image selection

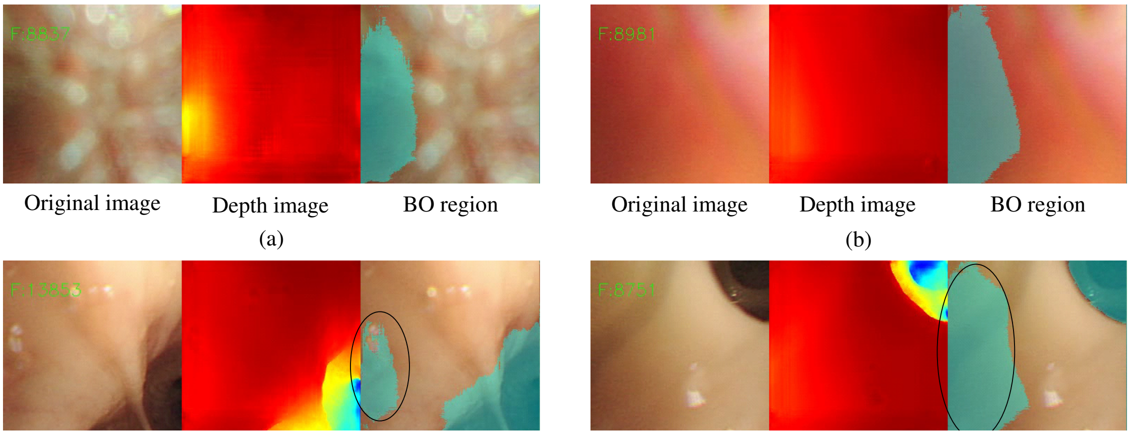


Figure 3.11: Examples of images showing poor results. In image (a) and (b) no BO region are observed; in (c) and (d), BO regions are mis-segmented (black ellipse).

procedure. For example, a CNN model is used to find an 'uninformative frame' that is either too dark or too blurry, etc [111].

The proposed method uses depth images estimated from CycleGAN. However, in some images, the depth image is not correctly estimated. This leads to the mis-segmentation of the BO region. Examples are shown in Figs. 3.11 (c) and (d). The region in the black ellipse is obviously mis-segmented. Therefore, an improvement of CycleGAN structure and create more training data may improve the accuracy of the estimated depth image.

3.7 Summary

This paper presented a BO segmentation method by using depth images obtained from CycleGAN. The proposed segmentation method showed promising results. Future work includes application of the segmentation result in a bronchoscopy navigation system and further improvement of the segmentation procedure.

Chapter 4

Deep learning-based branching level localization

This chapter describes the author's work on how to use the extracted anatomical structure to estimate the level of the branch. In the previous chapter, the depth image estimation method is introduced, this chapter describes our work of using depth-image to estimate the branching level for bronchoscopy navigation since it is vital to identify the examining branches in the bronchus tree during examinations.

4.1 Introduction

As described in the Introduction Chapter, a bronchoscope navigation system (BNS) [78, 81, 86] is used during bronchoscopy and transbronchial lung biopsies (TBLB) to assist bronchoscopists. Current BNSs rely on precise bronchoscope tracking results, i.e., these methods estimate the camera pose precisely for each frame. The precise tracking methods can be roughly divided into conventional and novel methods. Conventional methods include video-CT-based methods [78, 81, 86, 108], additional sensor-based methods [84, 93, 103, 104], and a hybrid of the two method types [94, 95]. Video-CT-based tracking estimates the camera pose via image registration. The camera pose is obtained from the registration between RB and virtual bronchoscopic (VB) images. In recent years this type has been extensively researched. Deguchi et al. improved the registration function by using anatomical structure [86]. Shen et al. used depth images to perform RB-VB registration and argued that depth images perform more robustly against illumination or bubbles than color images [98]. However, image similarity-based tracking is time-consuming, and the difference in RB and VB images leads to tracking error. Additional sensor-based bronchoscope tracking uses the output of an electromagnetic (EM) sensor to calculate the camera pose [84, 93, 103, 104], which is widely used in clinical applications [82, 84]. Hybrid tracking fuses the output of video-CT-based and additional sensor-based tracking [94, 95]. However, EM-sensor-based tracking is easily affected by metallic surgical tools, and the terminal bronchi cannot be examined due to the excessive size of the bronchoscope tip. Recently, pure RB image-based [130, 167] and deep learning-based [108, 109] bronchoscope tracking are becoming widely used. Pure RB image-based tracking (e.g., simultaneous localization and mapping) is used to estimate the camera's pose from the RB image by minimizing the reprojection error of 3D points in camera frames [130, 167]. The application of deep learning for tracking is still being explored. For example, it can be used to improve conventional tracking schemes [108] or to train a positioning network to locate RB images in CT images [109].

Unfortunately, precise bronchoscope tracking-based navigation fails easily due to the accumulated tracking error. Therefore, a bronchial branch localization-based bronchoscopy navigation scheme [155, 168] is proposed to overcome the shortcomings of conventional precise bronchoscope tracking methods. Since the bronchus has a tree-like structure, providing the depth of the branching level from the trachea is the core of this type of navigation system. Gil Debora et al. developed a branch level by loacating the bronchoscope in bronchus tree [155]. They used color image to extract the BO and generate the navigational information by proviing the anatomical structure of the bronchi. Shinohara et al. developed a branch identifying system that provides the anatomical name of the branch under examination [168]. The branching name is estimated by using RB image and pre-generated VB images in their eigenspace. Unlike conventional precise bronchoscope tracking-based navigation systems, this type of system localizes the bronchoscope in the bronchial branches. We plan to use only color RB images to build a navigation system rather than using both CT and RB images. We estimate the current branching level by integrating the number of BOs and the camera-moving direction. The number of BOs is estimated from depth images rather than color RB images. This is because a color image-based system [169] behaves poorly if there are bad imaging conditions such as strong illumination or bubbles [98]. The camera-moving direction is calculated using the results of feature point-based camera-motion estimation. These information is used to decide the current branching level.

4.2 Purpose of this Chapter

In this Chapter, a branching level estimation method is proposed based on the changes of the BO region and the camera moving direction. The BO regions are segmented from the depth images, which are estimated from CycleGAN by using bronchoscopic video. The camera moving direction is estimated from the feature point-based camera motion estimation. The branching level is estimated by using the changes of the orifice region and the camera moving direction.

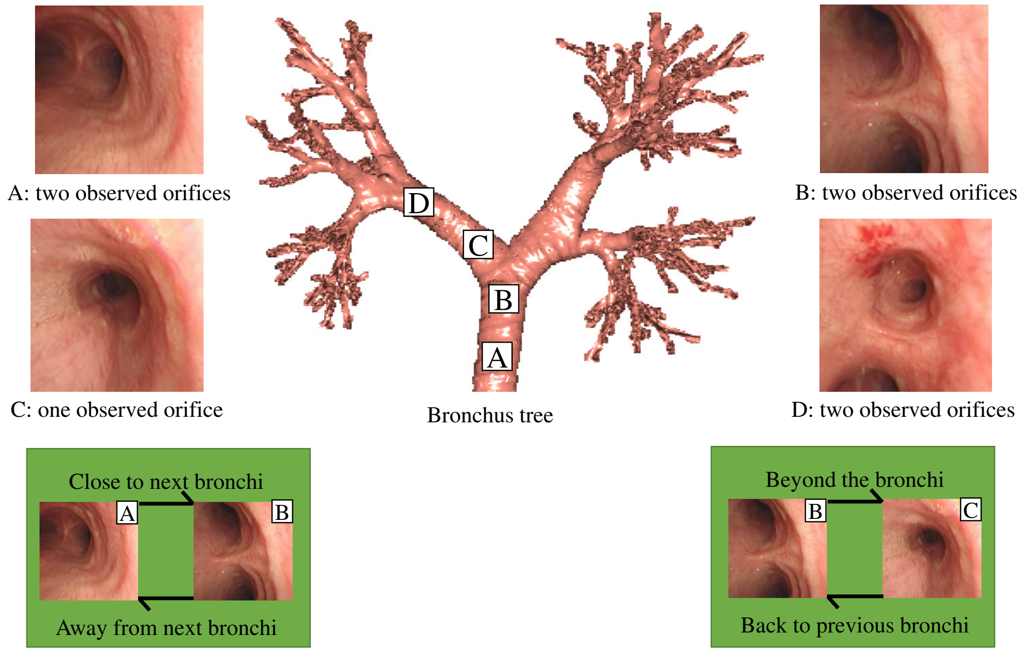


Figure 4.1: Example of the relationship between the BO and branching level. We select RB images in four locations to show the BO. Figures at the bottom show the branching level changes concerning the BO region and camera motion.

4.3 Proposed method

4.3.1 Overview

We estimate the current branching level, which represents the number of branches passed from the trachea, by using the changes in the BOs. The BO is one of the commonly observed anatomical structures in the bronchus scene, and its appearance in the RB image is related to the branches. We illustrate two possibilities for the BO's changes and the related changes of the branching level in Fig. 4.1 and summarize them as follows.

(1) The number of BOs remains the same (bottom-left figure). In this case, the camera stays at the same branching level regardless of the camera's moving direction.

Table 4.1: Possible outcomes for changes of BO number, camera-moving direction, and branching level

Change of BOs	Increase		Decrease		Same	
Camera-moving direction	*Forward*	*Backward*	*Forward*	*Backward*	*Forward*	*Backward*
Change of branching level	Same	Previous	Next	Same	Same	Same

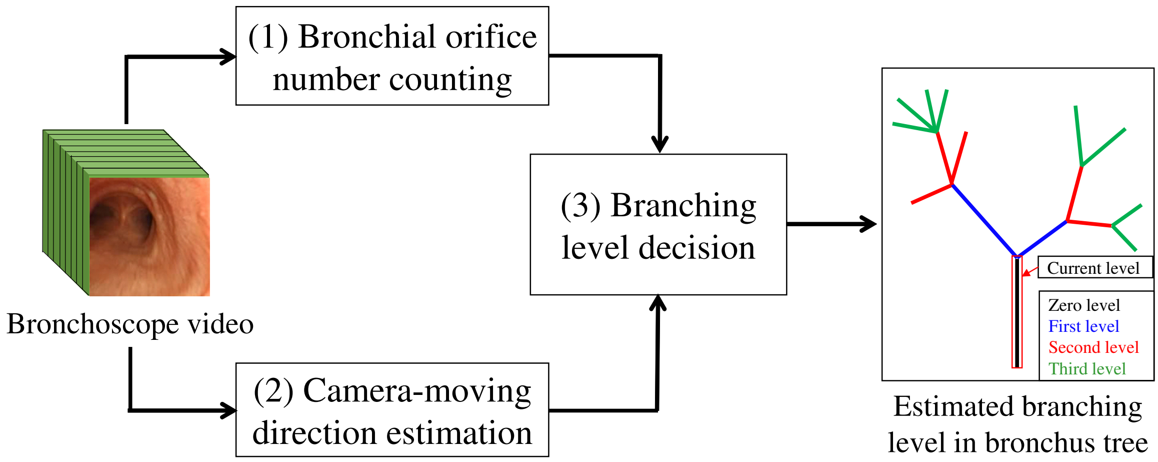


Figure 4.2: Flowchart of branching level estimation algorithm. RB video is used as input and branching level is output. There are three processes: counting the number of BOs, estimating the camera-moving direction, and estimating the branching level. The different branching levels are shown in different colors.

(2) The number of BO changes (bottom-right figure). In this case, the branching level changes as the number of BO changes. If the camera moves forward (e.g., from B to C), the branching level increases; if the camera moves backward (e.g., from C to B), the branching level decreases.

All possible relationships involving the BO's changes, camera motions, and branching level's changes are shown in Table 4.1.

We propose a branching level estimation method using only RB images. In the training phase, virtual depth images generated from CT images and RB images are used. The proposed branching level estimation algorithm is carried out using three processes (Fig. 4.2): (1) counting the BOs using depth images to determine the number of BOs in the RB image; (2) estimating the camera-moving direction using the results of feature point-based motion estimation; and (3) deciding the branching level by considering the changes in BOs and camera-moving direction.

4.3.2 Depth image-based BO counting

We used a depth image to identify the number of BOs instead of a color image because depth images behave more robustly in image-guided procedures [98]. We have used CycleGAN for depth iamge estimation in Chapter 3, since most of the depth image estimation procedure are same. We only introduce this estimation procedure briefly in this Chapter. The depth image-based BO counting algorithm is introduced from two perspectives: (1) CycleGAN for RB-to-depth translation; and (2) BO counting.

We use the prepared RB images and virtual depth images as the input of CycleGAN to find mapping X, represented by a trained generator. This generator is used to generate depth images by using RB images as input.

Depth image-based BO counting

The generated depth images are used to count the number of BOs. Since the BO region is farther from the camera position than the other regions, the BO region has a darker value in the depth image. An example is shown in Fig. 4.3: two BOs are observed in the RB image as well as in the depth image (red circles in Fig. 4.3). We count the number of BOs by counting the darker regions in the depth image.

The counting procedure works like the procedure of finding local minima described in Chapter 3. We vertically and horizontally project the image pixels and use their

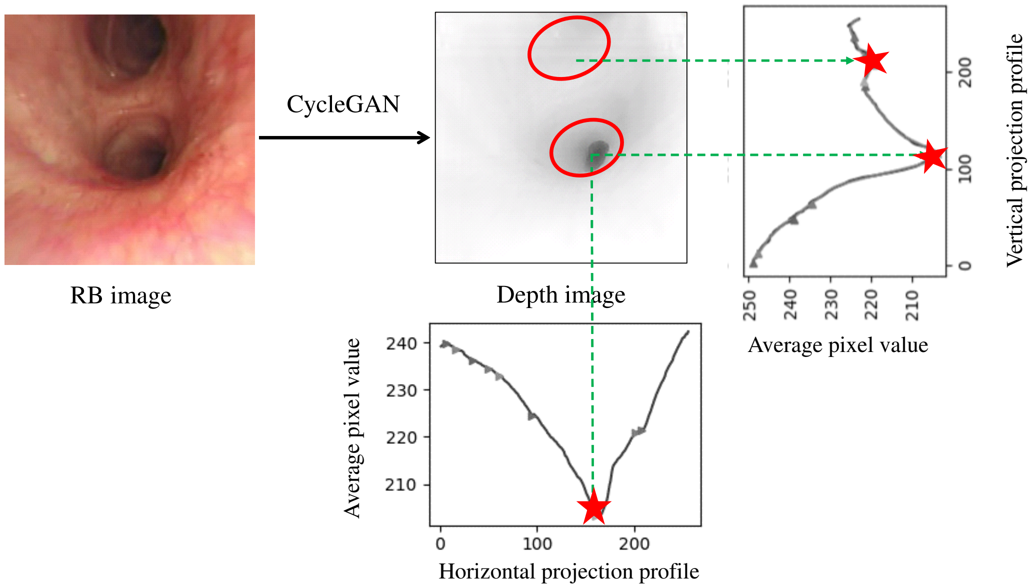


Figure 4.3: Examples of RB image, depth image, and vertical/horizontal projection profile of pixels in depth image. Red ovals approximately show BO region in depth image. Green dotted lines show projection from depth image in each direction. Darker region in depth image corresponds to wave trough in curve. We mark the desired local minimum values with stars. Other local minimum values are detected due to the influence of cartilage rings.

intensity profile of depth image **D** to find the number of darker regions. Here, we denote the pixel value at pixel (i, j) in a depth image as $\mathbf{D}(i, j)$. The i-th projected value in the vertical direction $I_x(i)$ is calculated using $I_x(i) = \sum_{j=1}^{H} \mathbf{D}(i, j)$, where H is the image height. The j-th projection value in the horizontal direction $I_y(j)$ is calculated using $I_y(j) = \sum_{i=1}^{W} \mathbf{D}(i, j)$, where W is the image width. As shown in Fig. 4.3, the darker region, which represents the BOs in the depth image, corresponds to the wave trough of curves $I_x(i)$ and $I_y(j)$ (stars in Fig. 4.3). Therefore, we calculate the local minima of the curves to obtain the number of BOs.

However, one different thing between the counting of BO and the segmentation of the BO is that the former procedure need more accurate calculation. Therefore, we perform a smoothing operation to accurately find the local minimum values: use downsampling and a mean filter to process the depth image; use polygon fitting [170] to fit two curves. We seek the local minimum values by using the derivative of $I_x(i)$ and $I_y(j)$. The local minima in $I_x(i)$ and $I_y(j)$ are denoted as N_x and N_y. The number of BOs is obtained to select the largest value from N_x and N_y. We describe $O^{(m)}$ as the number of BOs for the m-th frame. $O^{(m)}$ ranges from one to three in our experiment.

4.3.3 Estimation of camera-moving direction

The camera-moving direction is estimated based on the results of the feature point-based camera-motion estimation. It follows two steps: (1) camera-motion estimation and (2) the decision of camera-moving direction.

Camera-motion estimation

We estimate the camera motion between two frames based on the extracted feature points in frames. For two frames, the $(m-k)$-th frame and the m-th frame (k is a constant value), their camera transformation is estimated using three steps: (1) Find the corresponding feature points among two frames. We use **O**riented FAST and **R**otated **B**RIEF (ORB) [145] features to represent the extracted key points. We compare the feature points' location in the image and compute the Hamming distance of the descriptor of two feature points to find the correspondence in two frames. (2) Estimate the fundamental matrix using the paired feature points under a *RANSAC* scheme. (3) Recover the camera motion using the fundamental matrix. The camera motion between two frames is obtained by the decomposition of an essential matrix, which is calculated from the fundamental matrix and the camera's intrinsic parameters [171]. The translation vector $\mathbf{t} = (x, y, z)^{\mathrm{T}}$ (each axis is in $[-1, 1]$) represents the movements from the $(m-k)$-th frame to the m-th frame in three directions. It is used to determine the camera-moving direction.

Determination of camera motion

The camera-moving direction $M^{(m)}$ from the $(m-k)$-th frame to the m-th frame is determined using

$$M^{(m)} = \begin{cases} forward & h < z < 1 \\ backward & -1 \leq z \leq -h \\ others & -h < z < h, \end{cases}$$

where h is a pre-set threshold. This equation decides whether the camera-moving direction among two frames is *forward*, *backward*, or *others* (no obvious movement is found).

4.3.4 Branching level estimation

We list the possible changes of branching level with respect to the BO number's changes and the camera-moving direction in Table 4.1. The branching level is decided according to this table.

A state vector is used to represent the status of each frame, which is defined as $S^{(m)} = \left(O^{(m)}, \overline{M^{(m)}}, L^{(m)}\right)$ for the m-th frame, where $L^{(m)}$ is the current branching level. $O^{(m)}$ is estimated in Section 4.3.2, and $\overline{M^{(m)}}$ is calculated based on the camera-moving direction result from Section 4.3.3. $\overline{M^{(m)}}$ is the camera-moving direction of one frame using its previous neighboring frames. We use the direction that appears most often in the previous five frames to calculate $\overline{M^{(m)}}$. The current branching level $L^{(m)}$ is decided by considering the history status, $\overline{M^{(m)}}$ and $O^{(m)}$. The branching level is estimated according to Table 4.1: $O^{(m)}$ equals $O^{(m-1)}$, $O^{(m)}$ is larger than $O^{(m-1)}$, and $O^{(m)}$ is smaller than $O^{(m-1)}$.

(1) When $O^{(m)}$ equals $O^{(m-1)}$, the bronchoscope is in the same branching level.

(2) When $O^{(m)}$ is larger than $O^{(m-1)}$, there are two possibilities according to $\overline{M^{(m)}}$: if $\overline{M^{(m)}}$ = *forward*, the camera stays in the same bronchus; if $\overline{M^{(m)}}$ = *backward*, the camera returns to the previous branch.

(3) When $O^{(m)}$ is smaller than $O^{(m-1)}$, there are two possibilities according to $\overline{M^{(m)}}$: if $\overline{M^{(m)}}$ = *forward*, the camera moves to the next branch; if $\overline{M^{(m)}}$ = *backward*, the camera stays in the same branch.

The state vector is updated for each frame. The branching level, which is initialized before the estimation starts, is manually decided by the bronchoscopist. When the estimation is started in the trachea, we initialize the branching level to 0.

4.4 Experiment setting and results

We performed branch-level estimation using the proposed method. We used virtual depth images and color RB images in the experiment. The RB videos were obtained from different patients during bronchoscopy by physicians at two hospitals. The *in-vivo* RB images were taken using two types of bronchoscope (BF-240 and BF-260, Olympus, Tokyo), and the CT images were taken using a CT scanner (XVision, Toshiba Medical Systems, Tokyo). The phantom RB images used for training were taken using a bronchoscope (BF-200, Olympus, Tokyo) and the CT images were taken using a CT scanner (Aquilion 16, Toshiba Medical Systems, Tokyo). The original size of the RB image was 410×370 pixels. These CT images were taken several days before bronchoscopy. Information on the CT images was shown in Table 4.2. They were used to generate virtual depth images. The image size of the virtual depth image was 256×256 pixels.

We implemented our proposed method on a Ubuntu system with an Intel Xeon Gold 6134 processor (8 cores at 3.2 GHz). CycleGAN was trained on an NVIDIA Tesla V100 GPU. We used three *in-vivo* cases and a phantom case to prepare the training dataset.

Table 4.2: Acquisition parameters of CT images

Case	Type	Slice size (pixels)	Pixel size (mm)	Number of slices	Slice spacing (mm)	Thickness (mm)
1	*In-vivo*	512 × 512	0.63 × 0.63	183	1.00	5.00
2	*In-vivo*	512 × 512	0.63 × 0.63	76	2.00	2.00
3	*In-vivo*	512 × 512	0.63 × 0.63	195	1.00	2.00
4	Phantom	512 × 512	0.39 × 0.39	667	0.30	0.50

Table 4.3: Parameter settings in experiment (2)

CycleGAN training				Moving direction estimation	
Image size (pixels)	Optimizer	Epoch	Batch	k (frame)	h
256 × 256	Adam	200	10	2	0.1

The numbers of RB images were 1050, 324, 1015 in *in-vivo* cases and 742 in the phantom case; the numbers of virtual depth images were 732, 595, 1044 in *in-vivo* cases and 968 in the phantom case. We used another four cases for validation. The parameters of the experiment are shown in Tables 4.3 and 4.4. The networks were trained in 200 epochs with 10 mini-batches. The size of all images for training was set to 256 × 256 pixels, which was decided according to the literature [163]. We found that the depth image generated within 200 epochs is sufficient for BO-counting tasks, although more epochs are useful for finding better results. We performed image domain translation using the found generator. The generated depth images and the RB images were used to estimate the branching level. The parameters used in the smoothing procedure and camera-moving direction estimation were decided by using the training data of an *in-vivo* case. The parameters in the smoothing procedure were manually decided by comparing the number of BOs from smoothed curves and the manually created ground truth of the BO number; the parameters are manually adjusted by comparing the obtained camera-moving direction to the ground truth.

Table 4.4: Parameter settings in experiment (2)

Smoothing procedure	
Down-sampled image size (pixels)	Kernel of mean filter
128×128	(5, 5)

4.4.1 Evaluation of depth image-based BO counting

We evaluated the BO-counting accuracy by comparing the estimated number of BOs with the manually counted BOs in each image. The algorithm's accuracy, which is defined as the percentage of correctly counted frames among the total processed frames, was shown in Table 4.5. The average accuracy in all cases was 89.1 %.

4.4.2 Evaluation of branching level estimation

We evaluated the branching level estimation method using four *in-vivo* videos. We manually created a ground truth of the branching level for each frame. The algorithm's accuracy is defined as the percentage of continuously correctly counted frames (from the first frame to the frame whose branching level is not correctly estimated) among all processed frames. The proposed method's accuracy in the four cases is shown in Table 4.5. The average accuracy of the branching level estimation was 87.6 %. Screenshots of several results are shown in Figs. 4.4 and 4.5. We showed the RB image, the generated depth image (in a heat map), and the visualization of the estimated branching level in the bronchus tree.

We investigated the processing time per frame of the proposed method in four cases containing 2765 frames. The processing time for one frame includes loading the image to memory, estimating the depth image, counting the BOs, estimating the camera-moving direction and the branching level, and visualizing the results. The average processing time was about 61 ms per frame, including 47 ms for the camera-moving

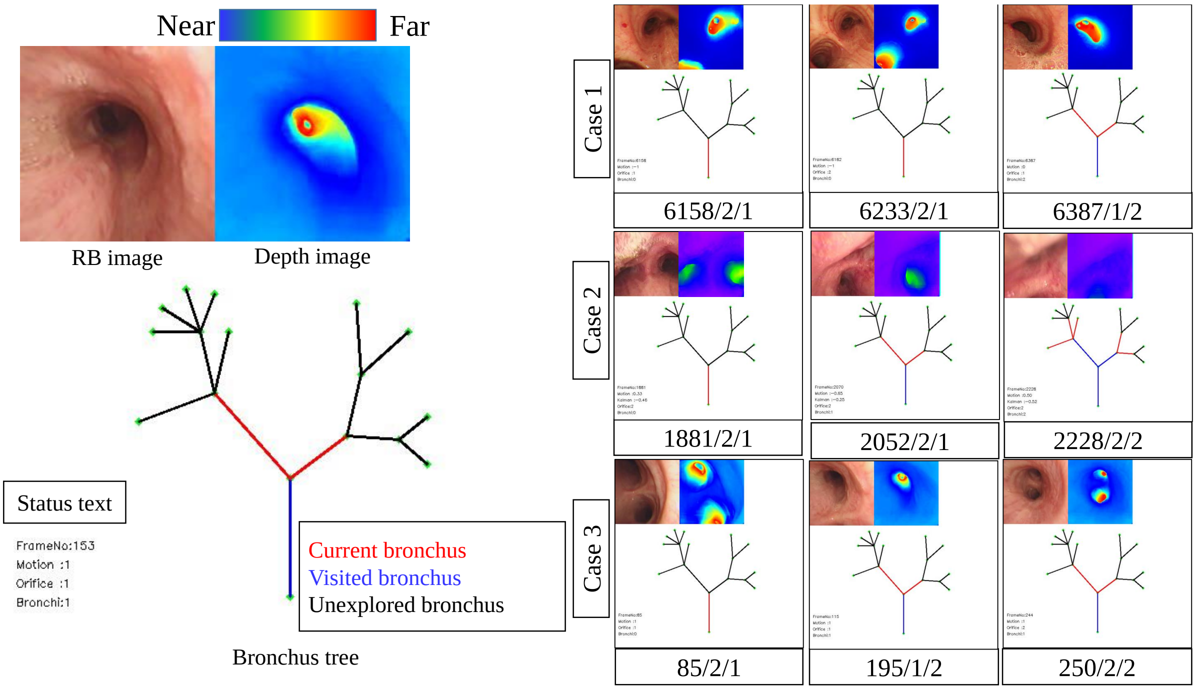
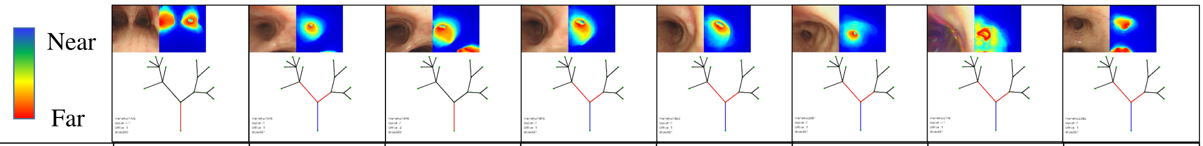


Figure 4.4: Example of estimated branching level in different cases: Left figure shows an example of entire interface; right figure shows several snapshots at different frames in different cases. Numbers below each snapshot denote frame number, number of estimated BOs, and estimated branching level. (For example, 6158/2/1 indicates that the current frame number is 6158, the detected BO number is two, and the current branching level is one.) Current branching level is shown in red, and visited levels are shown in blue. RB and depth images (visualized in heat map) are shown at top for quality evaluation.

Table 4.5: Accuracy of proposed method for branching level estimation

Case no.	No. of images	Accuracy of bronchial BO number counting	Accuracy of branching level estimation
1	990	88.3 %	99.9 %
2	1015	95.2 %	83.7 %
3	290	77.0 %	69.6 %
4	470	95.7 %	97.0 %
Average	691	89.1 %	87.6 %



Frame no.	1405		1540		1646		1819		1853		2061		2140		2382	
	P	GT	P	GT	P	GT	P	GT	P	GT	P	GT	P	GT	P	GT
No. of O	2	2	1	1	2	2	1	1	1	1	1	1	1	1	2	2
No. of L	1	1	2	2	1	1	2	2	2	2	2	2	1	1	2	2

Figure 4.5: Results of proposed method in *in-vivo*: O and L denote number of BO regions and estimated branching levels. P means estimated value from algorithms, and GT means the manually created ground truth. In this case, the moving trajectory of the bronchoscope is trachea $\rightarrow$ left main bronchus $\rightarrow$ trachea $\rightarrow$ left main bronchus.

direction estimation (CPU), 5 ms for the depth information prediction (GPU), 3 ms for counting the BOs (CPU), and 6 ms for estimation of the branching level (CPU).

4.5 Discussion

As shown in Table 4.5, average accuracy in the proposed branching level estimation method is 87.6 %. The estimation results in Figs. 4.4 and 4.5 show that the branching level is correctly estimated. We investigated the performance of several image-guided navigation schemes. Previous works [78, 81, 86, 108] estimated the precise camera position from RB images for navigation. Through decades of development, bronchoscope tracking has achieved high accuracy (e.g., 3.0 mm [167], 2.4 mm [109], and 3.18 mm [108]). On the other hand, we estimate the location of the bronchoscope in branching level for branch localization-based navigation. The two navigation types are different, and thus a direct comparison cannot be made. Our work is similar to a navigation scheme described previously [155]. We developed a new type of branching level navigation scheme using depth images rather than color images. Some of the advantages and limitations of the proposed method are outlined in the following section.

4.5.1 Advantages of proposed method

Simplified patient-CT registration

One advantage of the proposed method is the simplicity with which the estimated branching level can be registered to bronchus trees. Conventional navigation systems using bronchoscope tracking require precise patient-CT registration ([86, 103, 167]). The proposed method requires no complex registration procedure.

No accumulation of tracking errors

Another advantage of the proposed method is that no tracking errors are accumulated during processing. The proposed method estimates the passed branching level from the trachea. Since the branching level changes by an integer as the branch changes, the accumulation of tracking errors can be prevented. Therefore, the proposed method tracks more frames than does the previous precise bronchoscope-tracking navigation.

Handling scenes with complex imaging conditions

The proposed method shows robustness against various imaging conditions such as bubbles/froth and strong illumination. We think this is because the depth image performs better than the color image in the BO-counting process. As shown in the 1881^{st} frame in Case 2 of Fig. 4.4, the estimated branching level is correct even though bubbles are observed.

Applicability of data from multiple bronchoscopes

The proposed method uses depth images for the BO-counting task. Since variations in RB images greatly decreased in the depth images, our method performs well with different types of devices. We assume that the conventional bronchoscope images are used as input. Furthermore, the proposed method works well for images captured by different devices of the same manufacturer (Olympus, Tokyo). In our future work, we will test images obtained by other bronchoscopes of different manufacturers for validation.

4.5.2 Investigation of failed frames

We investigated the images that failed to estimate the branching levels. The first reason is the miscounting of BOs. The BO-counting method was robust in most scenes, but the jitter in the curves was not entirely eliminated. The second reason is that when the bronchoscope hits the bronchial wall, the BO counting fails. We show an example where branching level estimation fails in Fig. 4.6. The RB images in this example are taken from Case 3 when the estimation of branching level goes wrong. This 6195^{th} RB image shows that the bronchoscope nearly hit the bronchial wall, which leads to the failure of BO counting (since there are no local minima in the vertical and horizontal projection profiles). In the frames before this frame, one BO is observed and the camera-moving direction is *forward*. According to the algorithm, the branching level increases by one in this situation. However, the camera actually stays at the same branching level. To decrease the failures from miscounting BOs, an image classification procedure may be used to detect and remove images of poor quality.

4.5.3 Investigation of smoothing procedure in BO counting

We investigated the contributions of each step in the smoothing procedure by checking the BO's number after each smoothing step. We show one example of the detected local minima after each smoothing step in the vertical projection profile in Fig. 4.7. The detected local minima were marked in each smoothing step using a star. The detected number of BOs was nine in the vertical projection of the original depth image (corresponding to (b)), six after the down-sampling (corresponding to (c)), three after the image blur and down-sampling (corresponding to (d)), and two after the final step (corresponding to (e)). If the depth image is used directly, some values of the local minima are miscount as the number of BOs; however, they are from the cartilage ring (corresponding to (b) in Fig. 4.7). After the down-sampling procedure is applied, these miscounted values are eliminated (corresponding to (c) in Fig. 4.7). After the down-sampling and image blur procedures are applied, the detected local minima decreased (corresponding to (d) in Fig. 4.7). Finally, two local minima are correctly detected after these smoothing steps. We consider that the down-sampling step and the polygon-fitting step make more contributions to the smoothing procedure.

4.5.4 Future work

In this paper, we described our research on the estimation of branching levels based on the changes of BOs. The proposed method is still in the preliminary stage so it cannot be used in clinical application. According to physicians' comments, bronchoscopy will be of great benefit if the name of the bronchus is provided. In addition, if BO shape information (e.g., diameter of BO) is provided, the diagnosis of bronchial disease (e.g., bronchiectasis) will be easier. Also, it would be more helpful if a navigation system integrated the diagnostic report-generation function, which would automatically record the locations of the abnormal tissues and branches passed during bronchoscopy. There-

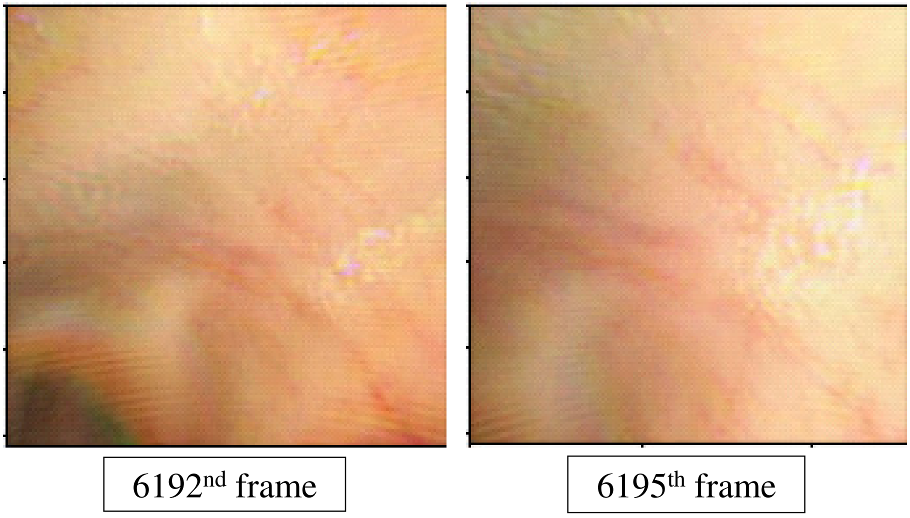


Figure 4.6: Examples of RB images that failed in branching level estimation. The number of BOs in the 6192^{nd} frame is one, while the number of BOs in the 6195^{th} frame is unknown. This is because the camera nearly hit the bronchial wall, so we cannot obtain useful information.

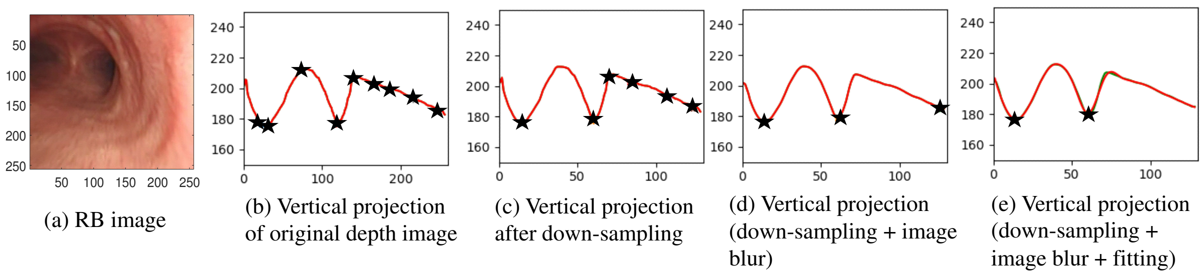


Figure 4.7: Example of the vertical projection profile and found local minima after each smoothing operation. The found local minima are marked using stars on the curves. The number of local minimum values are nine for (b), six for (c), three for (d), and two for (e). There are two BOs in this RB image.

fore, additional functions should be implemented before clinical application. We will utilize the estimated depth image to segment the BO region and use it for branch name estimation and the diagnosis of the bronchial disease. The navigation system we want to develop will provide navigational information, such as the name of the branch, in real-time to physicians.

4.6 Summary

This Chapter describes a new type of patient-specific bronchoscopy navigation scheme that uses the anatomical structure to estimate the branching level. This method only uses color RB images to estimate the branching level instead of traditional video-CT image registration based or additional sensor-based bronchoscope tracking. The proposed method behaves robustly to scenes showing poor imaging conditions such as strong illumination changes and bubbles. Since this method is still in its preliminary stage, the performance is still need to be improved. However, the proposed method provides a new approach for bronchoscopy navigation. In the future, more functions are expected to be implemented to make it a bronchoscopy navigation system.

Chapter 5

Conclusions and future work

This chapter briefly concludes the topics introduced in the previous Chapters at the beginning of the Chapter. Then discusses the advantages and disadvantages of the involved two tracking types. Finally, the promising directions of bronchoscope tracking in the future are discussed.

5.1 Summary

This thesis described the study on image-guided bronchoscope tracking for bronchoscopy navigation. More than 140 years ago, the endoscope application in bronchus examination made it possible to see the inside of the human bronchus. More than 100 years ago, the invention of X-ray photography made it possible to visualize the inside of the human body. As time goes by, more and more medical imaging techniques were invented and applied to clinical applications. These techniques have benefited bronchoscopy. According to the application, the diagnosis techniques are classified into computer-aided diagnosis (CAD) and computer-aided surgery (CAS). CAD procedure determines the presence or absence of lesions, and the decisions on treatment type are made based on the result of the CAD procedure. CAS procedure is used to assist physicians during surgery. This study aims at the CAS procedure by using RB images.

In Chapter 2, the pure RB image-based bronchoscope tracking is described. Bronchoscope tracking is an important component in bronchoscopy navigation. The previous video-CT-based method behaves weakly in regions lacking texture and suffers from the difference between preoperative VB and intraoperative RB images; additional sensor-based tracking can't enter the terminal bronchus due to the limitation of device size. Therefore, a pure RB image-based bronchoscope tracking is proposed using the visual SLAM technique in computer vision. This technique uses 3D-2D matches to estimate the camera pose and surroundings from a monocular video. To make it suitable for the bronchus scene, finding 3D-2D matches is improved with the stricter conditions. The experiment results in the phantom case show promising results of the proposed method. Moreover, the proposed method shows better performance in dealing with scenes of existing simulated deformation.

In Chapter 3, a deep-leaning-based anatomical structure segmentation method is described. The BO region is an important anatomic structure in the bronchus scene. Previous works use image appearance and geometry shape to extract the orifice region from RB images, which behaves poorly in scenes existing complex imaging condition. To overcome the shortcomings of the previous method, the orifice region is segmented using a depth image, which is estimated from CycleGAN. Experimental results showed that the proposed method has a higher DICE score than the previous methods comparing to ground truth; the proposed method segments each orifice region in complex imaging conditions.

In Chapter 4, a branching level estimation method is described. The conventional bronchoscope tracking method estimates the camera pose for each frame. Since the tracking procedure is based on the previous tracking results, the tracking error accumulates as the tracking of the new frames. To decrease the tracking error, a branching level estimation method is used. This method monitors the changes in the BO region. Once the orifice has changed among frames, the camera moving direction is used to decide the current branching level. This method uses the depth images estimated from CycleGAN to count the orifice region and shows good performance in poor imaging conditions. The proposed method shows the potential for bronchoscopy navigation.

5.2 Comparison of two tracking type

5.2.1 System complexity

Tracking complexity

The precise bronchoscope tracking-based systems precisely estimate each frame's camera pose, which could provide the pose of the bronchoscope camera with high accuracy. With the estimated camera pose and the registration result between CT images and RB images coordinate, the corresponding VB images could be generated and used for more augmented reality (AR) applications. However, on the other hand, this type of tracking is time-consuming and a waste of resources. Besides, CT images are required before the examination while in the case of emergency CT images are unavailable. Moreover, in several cases of bronchoscopy, there is no need to know the bronchoscope position in high accuracy. The precise bronchoscope tracking schemes, such as video-CT-based tracking and visual SLAM-based tracking, show good accuracy. Both of the tracking methods needs a huge amount of time for computation.

On the other side, the coarse bronchoscope localization only shows key navigational information, such as the branching level and anatomic structure changes during bronchoscopy. This type of navigation occupies fewer computer resources and provides essential information for navigation, which may be welcomed in the future.

Registration procedure

The registration procedure of the precise bronchoscope tracking-based systems needs a lot of effects to perform registration. The transformation in registration involves six or more degrees of freedom (DoF), including the translation in position and rotation in orientation. Once the unexpected scene happens, such as the patient movement, tracking error becoming larger, the registration procedure should be refined immediately to avoid further tracking error. However, in the clinical scene, these scenes easily tend to happen.

On the other hand, the coarse bronchoscope localization-based navigation system does not consider the camera pose in each frame. The registration is performed by considering the branching level. In the future, more functions are added to this system, which however, will not increase the complexity of registration procedure.

5.2.2 Tracking accuracy

The precise bronchoscope tracking-based navigation scheme has been researched for decades. Therefore, there are many efficient solutions aiming at solving different scenes. The accuracy of the CT-image registration-based schemes have achieved high accuracy (e.g., 3.0 mm [167], 2.4 mm [109], and 3.18 mm [108]). The tracking accuracy is quite high in these applications.

On the other hand, the coarse bronchoscope tracking method has not been so widely researched. Therefore, navigation systems based on this type of tracking have quite huge of potential applications clinically in the future.

5.2.3 Robustness

Since the conditions in RB image are very complex, the robustness of the navigation system is quite crucial. The conventional precise bronchoscope tracking-based navigation systems use color RB images, which may suffer from the difference between VB and RB images. The differences in RB and VB images will decrease tracking accuracy and finally lead to tracking failure.

The coarse bronchoscope localization-based navigation systems show robust in images existing poor imaging conditions. The proposed branching level estimation method is based on the depth image, in which the influence of the imaging conditions has been greatly decreased. Therefore, depth image-based tracking behaves more robustly. What's more, since the coarse bronchoscope localization-based scheme identifies the branching names, the tracking error of this system type will not accumulate like an exponential explosion.

5.3 Future work

This thesis describes two types of bronchoscopy navigation systems: precise bronchoscope tracking-based and coarse bronchoscope localization-based navigation system. In the following section, we will introduce the future work of two types of navigation systems respectively.

The precise bronchoscope tracking-based navigation has been developed for quite a long time. The currently widely used navigation system is EM sensor-based method navigation and in the future several years. However, in recent years, deep learning-based methods have been deeply researched. Several applications of the image-guided CAD systems have been developed and applied in clinical applications. The video-CT image registration-based bronchoscope tracking has many researches. In the future, the precise bronchoscope tracking-based navigation systems will become more robust and overcome more difficulties. The registration procedure will be easier to operate and the navigation system will be integrated with more functions. Moreover, the time cost of the whole system will be decreased. Physicians will use this type of navigation system more easily.

On the other hand, the bronchial branch localization-based navigation type still needs to get more attention to make it a clinical application. However, this type of navigation system has promising performance. As we mentioned in Chapter 4, if this type of navigation system could integrate more functions, physicians will be more willing to use it. In the future, this type of navigation system may become popular after more functions are integrated.

As conclusion, both of the conventional navigation system and popular schemes use the special anatomical structure of the bronchus and the existing prior knowledge about the bronchus. In the future, we think a specialised navigation system considering these points will have better performance in clinical applications and more tend to be welcomed.

Acknowledgement

Firstly, I would like to express my sincere gratitude to my supervisor, Prof. Dr. Kensaku Mori. Without his guidance, encouragement, and support, I could have never finished this study. He has been an excellent advisor and has always directed me in doing my fundamental research that will make a difference. Moreover, he has a deep understanding of professional knowledge, and a meticulous attitude in academic research. He helps me a lot not only in academic research but also in daily life.

I would also like to convey my heartfelt thanks to Prof. Dr. Masahiro Oda for his help during my doctoral study. The care and greetings from him have increased my confidence. Without his kind guidance, I could not finish my doctoral study smoothly. I am also very grateful to Prof. Dr. Katashi Nagao, Prof. Dr. Hiroaki Kudou and Prof. Dr. Daisuke Deguchi who gave me precious comments and were also willing to review my thesis. They have made my thesis better.

I want to express my special appreciation to Prof. Dr. Yuichiro Hayashi and Prof. Dr. Takayuki Kitasaka (Aichi Institute of Technology). Their rich knowledge and experience regarding my work were a great assistance to resolve my various doubts in regards to my dissertation. Their wisdom thinking and kindness support have greatly decreased the difficulties during my study. I have benefited a lot from them.

I owe my sincere gratitude to Prof. Dr. Hirotoshi Honma (Sapporo-Kosei General Hospital), Prof. Dr. Hirotsugu Takabatake (Sapporo Minami-Sanjo Hospital), Prof. Dr.

Masaki Mori (Sapporo-Kosei General Hospital), and Prof. Dr. Hiroshi Natori (Keiwakai Nishioka Hospital) for their helpful medical advice, their encouragement and their valuable materials. The remote video conference every Tuesday's night has been an expectant time to discuss the questions in research.

I am deeply grateful to my seniors at Mori laboratory: Dr. Chenglong Wang and Dr. Hirohisa Oda. Their insightful comments and suggestions in regard to my study since the Autumn of 2015. I would also like to thank all my colleagues those who are undergraduate and those who have graduated from Mori laboratory. Studying with them have been a unforgettable memory. There are other valuable memories such as lab travel, the special events happened in Mori laboratory.

I owe a great deal to all the organizations, who generously sponsored my study, including Chinese and Japanese government; Ministry of Education, Culture, Sports, Science and Technology (MEXT); the grant from Murata Science Foundation; the Olympus Corporation and other organizations. Their support and help have made this research work easier.

Finally I would like to extend my indebtedness to my parents, my wife and other relationships. Thanks for their support, continuous understanding, warmness encouragement and sacrifice, throughout the whole study period.

Bibliography

[1] Rebecca L Siegel, Kimberly D Miller, and Ahmedin Jemal. Cancer statistics, 2020. *CA: a cancer journal for clinicians*, 70(1):7–30, 2020.

[2] National Cancer Institute. Cancer stat facts: Lung and bronchus cancer. `https://seer.cancer.gov/statfacts/html/lungb.html`. Accessed at Nov. 15, 2020.

[3] Olympus Corporation. The bronchus through the bronchivideoscope poster. `https://www.olympusprofed.com/pulm/bronchoscopy/1158/`. Accessed at Nov. 24, 2020.

[4] Your Practice Online. Bronchoscopy. `https://www.ypo.education/medical-tests/bronchoscopy-t312/video/`. Accessed at Nov. 24, 2020.

[5] Rebecca L Siegel, Kimberly D Miller, and Ahmedin Jemal. Cancer statistics, 2019. *CA: a cancer journal for clinicians*, 69(1):7–34, 2019.

[6] Kathleen A Cronin, Andrew J Lake, Susan Scott, Recinda L Sherman, Anne-Michelle Noone, Nadia Howlader, S Jane Henley, Robert N Anderson, Albert U Firth, Jiemin Ma, et al. Annual report to the nation on the status of cancer, part I: National cancer statistics. *Cancer*, 124(13):2785–2800, 2018.

[7] Rui-Mei Feng, Yi-Nan Zong, Su-Mei Cao, and Rui-Hua Xu. Current cancer situa-

tion in China: good or bad news from the 2018 global cancer statistics? *Cancer Communications*, 39(1):22, 2019.

[8] Keisuke Onoi, Yusuke Chihara, Junji Uchino, Takayuki Shimamoto, Yoshie Morimoto, Masahiro Iwasaku, Yoshiko Kaneko, Tadaaki Yamada, and Koichi Takayama. Immune checkpoint inhibitors for lung cancer treatment: A review. *Journal of clinical medicine*, 9(5):1362, 2020.

[9] Akbar Khanmohammadi, Ali Aghaie, Ensieh Vahedi, Ali Qazvini, Mostafa Ghanei, Abbas Afkhami, Ali Hajian, and Hasan Bagheri. Electrochemical biosensors for the detection of lung cancer biomarkers: A review. *Talanta*, 206:120251, 2020.

[10] National Cancer Center. Lung cancer: examination. `https://ganjoho.jp/public/cancer/lung/diagnosis.html`. Accessed at Dec. 15, 2020.

[11] Ting Zhou, Zhonghan Zhang, Fan Luo, Yuanyuan Zhao, Xue Hou, Tingting Liu, Kai Wang, Hongyun Zhao, Yan Huang, and Li Zhang. Comparison of first-line treatments for patients with extensive-stage small cell lung cancer: a systematic review and network meta-analysis. *JAMA network open*, 3(10):e2015748–e2015748, 2020.

[12] Bertram Eugene Warren. *X-ray Diffraction*. Courier Corporation, 1990.

[13] Thorsten M Buzug. Computed tomography. In *Springer Handbook of Medical Technology*, pages 311–342. Springer, 2011.

[14] James Mattson and Merrill Simon. *The pioneers of NMR and magnetic resonance in medicine: The story of MRI*. Bar-Ilan University Press, 1996.

[15] Takayuki Kitasaka, Kensaku Mori, Jun-ichi Hasegawa, and Jun-ichiro Toriwaki. A method for extraction of bronchus regions from 3d chest x-ray ct images by

analyzing structural features of the bronchus. *FORMA-TOKYO-*, 17(4):321–338, 2002.

[16] Heang-Ping Chan, Lubomir M Hadjiiski, and Ravi K Samala. Computer-aided diagnosis in the era of deep learning. *Medical physics*, 47(5):e218–e227, 2020.

[17] Hiroshi Fujita. AI-based computer-aided diagnosis (AI-CAD): the latest review to read first. *Radiological physics and technology*, 13(1):6–19, 2020.

[18] Juri Yanase and Evangelos Triantaphyllou. A systematic survey of computer-aided diagnosis in medicine: Past and present developments. *Expert Systems with Applications*, 138:112821, 2019.

[19] Maryellen Lissak Giger, Kunio Doi, and Heber MacMahon. Image feature analysis and computer-aided diagnosis in digital radiography. 3. automated detection of nodules in peripheral lung fields. *Medical Physics*, 15(2):158–166, 1988.

[20] Muzzamil Javaid, Moazzam Javid, Muhammad Zia Ur Rehman, and Syed Irtiza Ali Shah. A novel approach to cad system for the detection of lung nodules in ct images. *Computer methods and programs in biomedicine*, 135:125–139, 2016.

[21] Xin-Wei Xu, Kunio Doi, Takeshi Kobayashi, Heber MacMahon, and Maryellen L Giger. Development of an improved cad scheme for automated detection of lung nodules in digital chest images. *Medical Physics*, 24(9):1395–1403, 1997.

[22] Jun-Gang Li, Wen-Xin Li, Jing-Ying Xu, Xiao-Qing Cai, Rui-Li Liu, Yong-Jun Li, Qun-Fen Zhao, and Qing-Nuan Li. Comparative study of pathological lesions induced by multiwalled carbon nanotubes in lungs of mice by intratracheal instillation and inhalation. *Environmental Toxicology: An International Journal*, 22(4):415–421, 2007.

[23] Armin Ernst, Gerard A Silvestri, and David Johnstone. Interventional pulmonary procedures: guidelines from the american college of chest physicians. *Chest*, 123(5):1693–1694, 2003.

[24] Hirohisa Horinouchi, Fumihiro Asano, Kenichi Okubo, Yoshinori Okada, Yoshinobu Ohsaki, Yuko Komase, Toshinori Hashizume, Mitsutomo Kohno, and Motoi Aoe. Current status of diagnostic and therapeutic bronchoscopy in japan: 2016 national survey of bronchoscopy. *Respiratory investigation*, 57(3):238–244, 2019.

[25] Olympus Corporation. Bronchoscopy. `https://www.mayoclinic.org/tests-procedures/bronchoscopy/about/pac-20384746#:~:text=Common%20reasons%20for%20needing%20bronchoscopy,provide%20treatment%20for%20lung%20problems`. Accessed at Nov. 15, 2020.

[26] Atilla P Kiraly, James P Helferty, Eric A Hoffman, Geoffrey McLennan, and William E Higgins. Three-dimensional path planning for virtual bronchoscopy. *IEEE Transactions on Medical Imaging*, 23(11):1365–1379, 2004.

[27] Dilek Hacer ÇEŞME. Effectiveness of virtual bronchoscopy in evaluation of centrally located lung cancer. *Bezmiâlem Science*, 9(2):226, 2021.

[28] The Johns Hopkins University. Bronchoscopy. `https://www.hopkinsmedicine.org/health/treatment-tests-and-therapies/bronchoscopy#:~:text=Bronchoscopy%20is%20a%20procedure%20to,)%2C%20and%20into%20the%20airways`. Accessed at Dec. 25, 2020.

[29] Eladio Rodriguez-Diaz, Samer Kaanan, Christopher Vanley, Tauseef Qureshi, and Irving J Bigio. Towards optical spectroscopy-guided lung biopsy: demonstration of tissue-type classification. *Journal of Biophotonics*, page e202100132, 2021.

[30] Tanmay S Panchabhai and Atul C Mehta. Historical perspectives of bronchoscopy. connecting the dots. *Annals of the American Thoracic Society*, 12(5):631–641, 2015.

[31] Francesco Petrella, Alessandro Borri, Monica Casiraghi, Sergio Cavaliere, Stefano Donghi, Domenico Galetta, Roberto Gasparri, Juliana Guarize, Alessandro Pardolesi, Piergiorgio Solli, et al. Operative rigid bronchoscopy: indications, basic techniques and results. *Multimed Man Cardiothorac Surg*, 2014:1–6, 2014.

[32] HD Becker and BR Marsh. Interventional bronchoscopy. *Anonymous. History of the rigid bronchoscope. Karger Publishers*, pages 2–15, 2000.

[33] Rainer ME Engel. Philipp bozzini—the father of endoscopy. *Journal of Endourology*, 17(10):859–862, 2003.

[34] Teruomi Miyazawa. History of the flexible bronchoscope. *Interventional bronchoscopy*, 30:16–21, 2000.

[35] HD Becker. Endobronchial ultrasound-a new perspective in bronchoscopy. *Lung cancer*, 1(16):112–113, 1996.

[36] Andrew RL Medford, JA Bennett, CM Free, and S Agrawal. Endobronchial ultrasound guided transbronchial needle aspiration. *Postgraduate medical journal*, 86(1012):106–115, 2010.

[37] Felix JF Herth, Ralf Eberhardt, Peter Vilmann, Mark Krasnik, and Armin Ernst. Real-time endobronchial ultrasound guided transbronchial needle aspiration for sampling mediastinal lymph nodes. *Thorax*, 61(9):795–798, 2006.

[38] Xiaonan Zang, Ronnarit Cheirsilp, Patrick D Byrnes, Trevor K Kuhlengel, Catherine Abendroth, Thomas Allen, Rickhesvar Mahraj, Jennifer Toth, Rebecca Bas-

com, and William E Higgins. Image-guided ebus bronchoscopy system for lung-cancer staging. *Informatics in Medicine Unlocked*, page 100665, 2021.

[39] Daniel P Steinfort, Felix JF Herth, Louis B Irving, and Phan T Nguyen. Safe performance of diagnostic bronchoscopy/ebus during the sars-cov-2 pandemic. *Respirology*, 25(7):703–708, 2020.

[40] Sebastian F Ameriso, Pierre Amarenco, Lesly A Pearce, Kanjana S Perera, George Ntaios, Wilfried Lang, Daniel Bereczki, Shinichiro Uchiyama, Scott E Kasner, Byung-Woo Yoon, et al. Intracranial and systemic atherosclerosis in the navigate esus trial: Recurrent stroke risk and response to antithrombotic therapy. *Journal of Stroke and Cerebrovascular Diseases*, 29(8):104936, 2020.

[41] Anushirvan Minokadeh and William C. Wilson. Chapter 49 - emergency airway management. In Allen Jeremias and David L. Brown, editors, *Cardiac Intensive Care (Second Edition)*, pages 598 – 631. W.B. Saunders, Philadelphia, second edition edition, 2010.

[42] Ko Pen Wang, Atul C Mehta, J Francis Turner, KoPen Wang, and Ko Pen Wang. *Flexible bronchoscopy*. Wiley Online Library, 2012.

[43] Bronchoscopy procedure. `https://www.jerseyshoreuniversitymedicalcenter.com/services/gastroenterology/bronchoscopy-procedure/#:~:text=The%20benefits%20of%20bronchoscopy%20include,can%20be%20removed%20from%20lungs`. Accessed at Dec. 15, 2020.

[44] Pyng Lee. Indications and limitations of bronchoscopy. In *Journal of Thoracic Oncology*, volume 2, pages S278–S281, 2007.

[45] Physician P.C. Health & Breathing Center MXBowen. The advan-

tages of navigational bronchoscopy. `https://www.mxbowenppc.com/2016/01/the-advantages-of-navigational-bronchoscopy/`. Accessed at Nov. 24, 2020.

[46] Stefan Milz, Georg Arbeiter, Christian Witt, Bassam Abdallah, and Senthil Yogamani. Visual SLAM for automated driving: Exploring the applications of deep learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 247–257, 2018.

[47] Md Mahmud Hasan, Beibut Amirgaliyev, Dastan Rakhatov, Almas Tuyakbayev, and Chingiz Kenshimov. Real time road mapping for the driving safety purposes. In *2015 International Conference on Connected Vehicles and Expo (ICCVE)*, pages 60–61. IEEE, 2015.

[48] Mike Zdeb. Driving distances and times using SAS® and google maps. In *SAS Global Forum*, volume 2010, 2010.

[49] Michelle Birdsall. Google and ite: The road ahead for self-driving cars. *Institute of Transportation Engineers. ITE Journal*, 84(5):36, 2014.

[50] Hao Yu, Shu Yang, Weihao Gu, and Shaoyu Zhang. Baidu driving dataset and end-to-end reactive control model. In *2017 IEEE Intelligent Vehicles Symposium (IV)*, pages 341–346. IEEE, 2017.

[51] Jun Liu, Jihua Xiao, HongJie Cao, and Jiakai Deng. The status and challenges of high precision map for automated driving. In *China Satellite Navigation Conference*, pages 266–276. Springer, 2019.

[52] Hongzhu Ruan, Xiaofeng Ji, and Chuan Feng. The empirical research on information behavior characteristics and satisfaction of drivers based on smart phone. *Procedia engineering*, 137:343–351, 2016.

[53] Martin J Citardi and Pete S Batra. Intraoperative surgical navigation for endoscopic sinus surgery: rationale and indications. *Current opinion in otolaryngology & head and neck surgery*, 15(1):23–27, 2007.

[54] Simon Leonard, Ayushi Sinha, Austin Reiter, Masaru Ishii, Gary L Gallia, Russell H Taylor, and Gregory D Hager. Evaluation and stability analysis of video-based navigation system for functional endoscopic sinus surgery on in vivo clinical data. *IEEE transactions on medical imaging*, 37(10):2185–2195, 2018.

[55] Alexander G Chiu and Winston C Vaughan. Revision endoscopic frontal sinus surgery with surgical navigation. *Otolaryngology–Head and Neck Surgery*, 130(3):312–318, 2004.

[56] Simon Leonard, Austin Reiter, Ayushi Sinha, Masaru Ishii, Russell H Taylor, and Gregory D Hager. Image-based navigation for functional endoscopic sinus surgery using structure from motion. In *Medical Imaging 2016: Image Processing*, volume 9784, page 97840V. International Society for Optics and Photonics, 2016.

[57] Anand Velusamy, Aishwarya Anand, and Nazrin Hameed. Navigation assisted frontal sinus osteoplastic flap surgeries–a case series. *Indian Journal of Otolaryngology and Head & Neck Surgery*, pages 1–5, 2021.

[58] Keiichi Tsuchida, Héctor M García-García, Willem J van der Giessen, Eugène P McFadden, Martin van der Ent, Georgios Sianos, Hans Meulenbrug, Andrew TL Ong, and Patrick W Serruys. Guidewire navigation in coronary artery stenoses using a novel magnetic navigation system: first clinical experience. *Catheterization and cardiovascular interventions*, 67(3):356–363, 2006.

[59] Christopher W Roy, John Heerfordt, Davide Piccini, Giulia Rossi, Anna Giulia

Pavon, Juerg Schwitter, and Matthias Stuber. Motion compensated whole-heart coronary cardiovascular magnetic resonance angiography using focused navigation (fNAV). *Journal of Cardiovascular Magnetic Resonance*, 23(1):1–17, 2021.

[60] Jun-tan Li, Xiang Gao, and Xu Li. Comparison of iassist navigation system with conventional techniques in total knee arthroplasty: A systematic review and meta-analysis of radiographic and clinical outcomes. *Orthopaedic surgery*, 11(6):985–993, 2019.

[61] Dong-Yeong Lee, Young-Jin Park, Sun-Chul Hwang, Jin-Sung Park, and Dong-Geun Kang. No differences in mid-to long-term outcomes of computer-assisted navigation versus conventional total knee arthroplasty. *Knee Surgery, Sports Traumatology, Arthroscopy*, 28(10):3183–3192, 2020.

[62] Babar Kayani, Sujith Konan, Jenni Tahmassebi, Atif Ayuob, Peter D Moriarty, and Fares S Haddad. Robotic-arm assisted medial unicondylar knee arthroplasty versus jig-based unicompartmental knee arthroplasty with navigation control: study protocol for a prospective randomised controlled trial. *Trials*, 21(1):1–11, 2020.

[63] Zaid Shihab, Catriona Clayworth, and Naveen Nara. Handheld, accelerometer-based navigation versus conventional instrumentation in total knee arthroplasty: a meta-analysis. *ANZ Journal of Surgery*, 90(10):2068–2079, 2020.

[64] Tomoyoshi Okamoto, Shinji Onda, Katsuhiko Yanaga, Naoki Suzuki, and Asaki Hattori. Clinical application of navigation surgery using augmented reality in the abdominal field. *Surgery today*, 45(4):397–406, 2015.

[65] Veronica Penza, Domenico Soriero, Giacinto Barresi, Davide Pertile, Stefano Scabini, and Leonardo S Mattos. The GPS for surgery: A user-centered eval-

uation of a navigation system for laparoscopic surgery. *The International Journal of Medical Robotics and Computer Assisted Surgery*, 16(5):1–13, 2020.

[66] Takeshi Aoki, Doaa A Mansour, Tomotake Koizumi, Yusuke Wada, Yuta Enami, Akira Fujimori, Tomokazu Kusano, Kazuhiro Matsuda, Koji Nogaki, Yoshihiko Tashiro, et al. Laparoscopic liver surgery guided by virtual real-time ct-guided volume navigation. *Journal of Gastrointestinal Surgery*, pages 1–8, 2020.

[67] JY An, JS Min, H Hur, YJ Lee, GS Cho, YK Park, MR Jung, JH Park, WJ Hyung, SH Jeong, et al. Laparoscopic sentinel node navigation surgery versus laparoscopic gastrectomy with lymph node dissection for early gastric cancer: short-term outcomes of a multicentre randomized controlled trial (SENORITA). *Journal of British Surgery*, 107(11):1429–1439, 2020.

[68] Yuichiro Hayashi, Kazunari Misawa, Masahiro Oda, David J Hawkes, and Kensaku Mori. Clinical application of a surgical navigation system based on virtual laparoscopy in laparoscopic gastrectomy for gastric cancer. *International journal of computer assisted radiology and surgery*, 11(5):827–836, 2016.

[69] Maryellen L Giger and Kenji Suzuki. Computer-aided diagnosis. In *Biomedical information technology*, pages 359–XXII. Elsevier, 2008.

[70] Kunio Doi. Computer-aided diagnosis in medical imaging: historical review, current status and future potential. *Computerized medical imaging and graphics*, 31(4-5):198–211, 2007.

[71] Ayman El-Baz and Jasjit S Suri. *Lung imaging and computer aided diagnosis*. CRC Press, 2011.

[72] Robert A Drebin, Loren Carpenter, and Pat Hanrahan. Volume rendering. *ACM Siggraph Computer Graphics*, 22(4):65–74, 1988.

[73] Toriwaki Jun-ichiro Mori Kensaku, Suenaga Yasuhito. Fast volume rendering based on software optimisation using multimedia instructions on pc platforms. In *International Journal of Computer Assisted Radiology and Surgery (CARS)*, pages 467–472, 2002.

[74] Kensaku Mori, Yasuhito Suenaga, and Jun-ichiro Toriwaki. Fast software-based volume rendering using multimedia instructions on pc platforms and its application to virtual endoscopy. In *Medical Imaging 2003: Physiology and Function: Methods, Systems, and Applications*, volume 5031, pages 111–122. International Society for Optics and Photonics, 2003.

[75] Yehuda Schwarz, Joel Greif, Heinrich D Becker, Armin Ernst, and Atul Mehta. Real-time electromagnetic navigation bronchoscopy to peripheral lung lesions using overlaid ct images: the first human study. *Chest*, 129(4):988–994, 2006.

[76] Marco Feuerstein, Takamasa Sugiura, Daisuke Deguchi, Tobias Reichl, Takayuki Kitasaka, and Kensaku Mori. Marker-free registration for electromagnetic navigation bronchoscopy under respiratory motion. In *International Workshop on Medical Imaging and Virtual Reality*, pages 237–246. Springer, 2010.

[77] Kensaku Mori, Daisuke Deguchi, Takayuki Kitasaka, Yasuhito Suenaga, Yosihnori Hasegawa, Kazuyoshi Imaizumi, and Hirotsugu Takabatake. Improvement of accuracy of marker-free bronchoscope tracking using electromagnetic tracker based on bronchial branch information. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 535–542. Springer, 2008.

[78] Kensaku Mori, Daisuke Deguchi, Jun Sugiyama, Yasuhito Suenaga, Jun-ichiro Toriwaki, Calvin R Maurer Jr, Hirotsugu Takabatake, and Hiroshi Natori. Tracking of a bronchoscope using epipolar geometry analysis and intensity-based im-

age registration of real and virtual endoscopic images. *Medical Image Analysis*, 6(3):321–336, 2002.

[79] Parya Jafari, Sergio Dempsey, Douglas A Hoover, Elham Karami, Stewart Gaede, Ali Sadeghi-Naini, Ting Yim Lee, and Abbas Samani. In-vivo lung biomechanical modeling for effective tumor motion tracking in external beam radiation therapy. *Computers in Biology and Medicine*, 130:104231, 2021.

[80] Sujin Yoon, Tran Minh Tam, Prathish K Rajaraman, Ching-Long Lin, Merryn Tawhai, Eric A Hoffman, and Sanghun Choi. An integrated 1d breathing lung simulation with relative hysteresis of airway structure and regional pressure for healthy and asthmatic human lungs. *Journal of Applied Physiology*, 129(4):732–747, 2020.

[81] Ivan Bricault, Gilbert Ferretti, and Philippe Cinquin. Registration of real and ct-derived virtual bronchoscopic images to assist transbronchial biopsy. *IEEE transactions on medical imaging*, 17(5):703–714, 1998.

[82] Thomas R Gildea, Peter J Mazzone, Demet Karnak, Moulay Meziane, and Atul C Mehta. Electromagnetic navigation diagnostic bronchoscopy: a prospective study. *American journal of respiratory and critical care medicine*, 174(9):982–989, 2006.

[83] Ralf Eberhardt, Nicolas Kahn, Daniela Gompelmann, Maren Schumann, Claus Peter Heussel, and Felix JF Herth. Lungpoint—a new approach to peripheral lesions. *Journal of Thoracic Oncology*, 5(10):1559–1563, 2010.

[84] Liat Appelbaum, Jacob Sosna, Yizhak Nissenbaum, Alexander Benshtein, and S Nahum Goldberg. Electromagnetic navigation system for ct-guided biopsy of small lesions. *American Journal of Roentgenology*, 196(5):1194–1200, 2011.

[85] Takashi Ishida, Fumihiro Asano, Koichi Yamazaki, Naofumi Shinagawa, Satoshi Oizumi, Hiroshi Moriya, Mitsuru Munakata, Masaharu Nishimura, Virtual Navigation in Japan (V-NINJA) trial group, et al. Virtual bronchoscopic navigation combined with endobronchial ultrasound to diagnose small peripheral pulmonary lesions: a randomised trial. *Thorax*, 66(12):1072–1077, 2011.

[86] Daisuke Deguchi, Kensaku Mori, Marco Feuerstein, Takayuki Kitasaka, Calvin R Maurer Jr, Yasuhito Suenaga, Hirotsugu Takabatake, Masaki Mori, and Hiroshi Natori. Selective image similarity measure for bronchoscope tracking based on image registration. *Medical Image Analysis*, 13(4):621–633, 2009.

[87] Xióngbiāo Luó, Marco Feuerstein, Takayuki Kitasaka, and Kensaku Mori. Robust bronchoscope motion tracking using sequential monte carlo methods in navigated bronchoscopy: dynamic phantom and patient validation. *International journal of computer assisted radiology and surgery*, 7(3):371–387, 2012.

[88] Daisuke Deguchi, Kensaku Mori, Jun-ichi Hasegawa, Jun-ichiro Toriwaki, Hirotsugu Takabatake, and Hiroshi Natori. Camera motion tracking of real bronchoscope using epipolar geometry analysis and ct-derived bronchoscopic images. In *Medical Imaging 2002: Physiology and Function from Multidimensional Images*, volume 4683, pages 30–41. International Society for Optics and Photonics, 2002.

[89] 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.

[90] Mark R Bowling, Erik E Folch, Sandeep J Khandhar, Jordan Kazakov, William S Krimsky, Gregory P LeMense, Philip A Linden, Boris A Murillo, Michael A Nead, Michael A Pritchett, et al. Fiducial marker placement with electromagnetic navigation bronchoscopy: a subgroup analysis of the prospective, multicenter navi-

gate study. *Therapeutic advances in respiratory disease*, 13:1753466619841234, 2019.

[91] Xiongbiao Luo and Kensaku Mori. Robust endoscope motion estimation via an animated particle filter for electromagnetically navigated endoscopy. *IEEE Transactions on Biomedical Engineering*, 61(1):85–95, 2013.

[92] Kensaku Mori, Daisuke Deguchi, Kazuyoshi Ishitani, Takayuki Kitasaka, Yasuhito Suenaga, Yosihnori Hasegawa, Kazuyoshi Imaizumi, and Hirotsugu Takabatake. Bronchoscope tracking without fiducial markers using ultra-tiny electromagnetic tracking system and its evaluation in different environments. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 644–651. Springer, 2007.

[93] Shuenn-Wen Kuo, Ying-Fan Tseng, Kuan-Yu Dai, Yeun-Chung Chang, Ke-Cheng Chen, and Jang-Ming Lee. Electromagnetic navigation bronchoscopy localization versus percutaneous ct-guided localization for lung resection via video-assisted thoracoscopic surgery: a propensity-matched study. *Journal of clinical medicine*, 8(3):379, 2019.

[94] Kensaku Mori, Daisuke Deguchi, Kenta Akiyama, Takayuki Kitasaka, Calvin R Maurer, Yasuhito Suenaga, Hirotsugu Takabatake, Masaki Mori, and Hiroshi Natori. Hybrid bronchoscope tracking using a magnetic tracking sensor and image registration. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 543–550. Springer, 2005.

[95] Scott A Merritt, Rahul Khare, Rebecca Bascom, and William E Higgins. Interactive ct-video registration for the continuous guidance of bronchoscopy. *IEEE transactions on Medical Imaging*, 32(8):1376–1396, 2013.

[96] Jiro Nagao, Kensaku Mori, Tsutomu Enjouji, Daisuke Deguchi, Takayuki Kitasaka, Yasuhito Suenaga, Jun-ichi Hasegawa, Jun-ichiro Toriwaki, Hirotsugu Takabatake, and Hiroshi Natori. Fast and accurate bronchoscope tracking using image registration and motion prediction. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 551–558. Springer, 2004.

[97] Fani Deligianni, Adrian Chung, and Guang Zhong. Predictive camera tracking for bronchoscope simulation with condensation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 910–916. Springer, 2005.

[98] M. Shen, S. Giannarou, and G. Z. Yang. Robust camera localisation with depth reconstruction for bronchoscopic navigation. *Int J CARS*, 10(6):801–813, 2015.

[99] Mali Shen, Stamatia Giannarou, Pallav L Shah, and Guang-Zhong Yang. Branch: Bifurcation recognition for airway navigation based on structural characteristics. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 182–189. Springer, 2017.

[100] Marco Visentini-Scarzanella, Danail Stoyanov, and Guang-Zhong Yang. Metric depth recovery from monocular images using shape-from-shading and specularities. In *2012 19th IEEE International Conference on Image Processing*, pages 25–28. IEEE, 2012.

[101] Daisuke Deguchi, Kazuyoshi Ishitani, Takayuki Kitasaka, Kensaku Mori, Yasuhito Suenaga, Hirotsugu Takabatake, Masaki Mori, and Hiroshi Natori. A method for bronchoscope tracking using position sensor without fiducial markers. In *Medical Imaging 2007: Physiology, Function, and Structure from Medical Images*, volume 6511, page 65110N. International Society for Optics and Photonics, 2007.

[102] Duane C Cornish and William E Higgins. Bronchoscopy guidance system based on bronchoscope-motion measurements. In *Medical Imaging 2012: Image-Guided Procedures, Robotic Interventions, and Modeling*, volume 8316, page 83161G. International Society for Optics and Photonics, 2012.

[103] 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.

[104] Ingmar Wegner, Juergen Biederer, Ralf Tetzlaff, Ivo Wolf, and Hans-Peter Meinzer. Evaluation and extension of a navigation system for bronchoscopy inside human lungs. In *Medical Imaging 2007: Visualization and Image-Guided Procedures*, volume 6509, page 65091H. International Society for Optics and Photonics, 2007.

[105] Xiongbiao Luo and Kensaku Mori. Robust endoscope motion estimation via an animated particle filter for electromagnetically navigated endoscopy. *IEEE Transactions on Biomedical Engineering*, 61(1):85–95, 2013.

[106] Timothy D Soper, David R Haynor, Robb W Glenny, and Eric 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, 2009.

[107] Xiongbiao Luo, Marco Feuerstein, Takamasa Sugiura, Takayuki Kitasaka, Kazuyoshi Imaizumi, Yoshinori Hasegawa, and Kensaku Mori. Towards hybrid bronchoscope tracking under respiratory motion: evaluation on a dynamic motion phantom. In *Medical Imaging 2010: Visualization, Image-Guided Procedures,*

and Modeling, volume 7625, page 76251B. International Society for Optics and Photonics, 2010.

[108] Mali Shen, Yun Gu, Ning Liu, and Guang-Zhong Yang. Context-aware depth and pose estimation for bronchoscopic navigation. *IEEE Robotics and Automation Letters*, 4(2):732–739, 2019.

[109] Jake Sganga, David Eng, Chauncey Graetzel, and David Camarillo. Offsetnet: Deep learning for localization in the lung using rendered images. In *2019 International Conference on Robotics and Automation (ICRA)*, pages 5046–5052. IEEE, 2019.

[110] David Recasens, Jose Lamarca, Jose M Facil, JMM Montiel, and Javier Civera. Endo-depth-and-motion: Reconstruction and tracking in endoscopic videos using depth networks and photometric constraints. *IEEE Robotics and Automation Letters*, 2021.

[111] Matthew I McTaggart and William E Higgins. Robust video-frame classification for bronchoscopy. *SPIE: Medical Imaging 2019: Image-Guided Procedures, Robotic Interventions, and Modeling*, 10951:109511Q, 2019.

[112] Li-Jia Li, Richard Socher, and Li Fei-Fei. Towards total scene understanding: Classification, annotation and segmentation in an automatic framework. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 2036–2043. IEEE, 2009.

[113] Dana H Ballard and Christopher M Brown. Computer vision. englewood cliffs. *J: Prentice Hall*, 1982.

[114] Alex M Andrew. Multiple view geometry in computer vision. *Kybernetes*, 2001.

[115] Wikipedia. Computer vision. https://en.wikipedia.org/wiki/Computer_vision. Accessed at June 15, 2021.

[116] Marco Visentini-Scarzanella, Takamasa Sugiura, Toshimitsu Kaneko, and Shinichiro Koto. Deep monocular 3d reconstruction for assisted navigation in bronchoscopy. *International journal of computer assisted radiology and surgery*, 12(7):1089–1099, 2017.

[117] Arm Limited. Computer vision. https://www.arm.com/ja/glossary/computer-vision. Accessed at Dec. 14, 2020.

[118] Andrew J Davison. Real-time simultaneous localisation and mapping with a single camera. In *Proceedings Ninth IEEE International Conference on Computer Vision*, pages 1403–1411. IEEE, 2003.

[119] Oscar G Grasa, Ernesto Bernal, Santiago Casado, Ismael Gil, and JMM Montiel. Visual slam for handheld monocular endoscope. *IEEE transactions on medical imaging*, 33(1):135–146, 2013.

[120] Oscar G Grasa, J Civera, A Guemes, V Munoz, and JMM Montiel. EKF monocular SLAM 3D modeling, measuring and augmented reality from endoscope image sequences. In *Medical image computing and computer-assisted intervention (MICCAI)*, volume 2, 2009.

[121] W Clements Logan, C Chapman William, M Dawant Benoit, L Galloway Jr Robert, and I Miga Michael. Robust surface registration using salient anatomical features for image-guided liver surgery: Algorithm and validation. *Med. Phys.*, 35(6):2528–2540, 2008.

[122] Oscar G. Grasa, Javier Civera, and J. M M Montiel. EKF monocular SLAM with

relocalization for laparoscopic sequences. *Proc. - IEEE Int. Conf. Robot. Autom.*, pages 4816–4821, 2011.

[123] Danail Stoyanov, Marco Visentini Scarzanella, Philip Pratt, and Guang-Zhong Yang. Real-time stereo reconstruction in robotically assisted minimally invasive surgery. *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, 6361 LNCS(PART 1):275–282, 2010.

[124] Andrew J Davison, Ian D Reid, Nicholas D Molton, and Olivier Stasse. Monoslam: Real-time single camera slam. *IEEE transactions on pattern analysis and machine intelligence*, 29(6):1052–1067, 2007.

[125] Georg Klein and David Murray. Parallel tracking and mapping for small ar workspaces. In *2007 6th IEEE and ACM international symposium on mixed and augmented reality*, pages 225–234. IEEE, 2007.

[126] Richard A Newcombe, Steven J Lovegrove, and Andrew J Davison. Dtam: Dense tracking and mapping in real-time. In *2011 international conference on computer vision*, pages 2320–2327. IEEE, 2011.

[127] Christian Forster, Matia Pizzoli, and Davide Scaramuzza. Svo: Fast semi-direct monocular visual odometry. In *2014 IEEE international conference on robotics and automation (ICRA)*, pages 15–22. IEEE, 2014.

[128] Jakob Engel, Thomas Schöps, and Daniel Cremers. Lsd-slam: Large-scale direct monocular slam. In *European conference on computer vision*, pages 834–849. Springer, 2014.

[129] Jakob Engel, Vladlen Koltun, and Daniel Cremers. Direct sparse odometry. *IEEE transactions on pattern analysis and machine intelligence*, 40(3):611–625, 2017.

[130] Raul Mur-Artal, Jose Maria Martinez Montiel, and Juan D Tardos. Orb-slam: a versatile and accurate monocular slam system. *IEEE transactions on robotics*, 31(5):1147–1163, 2015.

[131] Raul Mur-Artal and Juan D Tardós. Orb-slam2: An open-source slam system for monocular, stereo, and rgb-d cameras. *IEEE Transactions on Robotics*, 33(5):1255–1262, 2017.

[132] Carlos Campos, Richard Elvira, Juan J Gómez Rodríguez, José MM Montiel, and Juan D Tardós. Orb-slam3: An accurate open-source library for visual, visual–inertial, and multimap slam. *IEEE Transactions on Robotics*, 2021.

[133] Ruihao Li, Sen Wang, and Dongbing Gu. Deepslam: A robust monocular slam system with unsupervised deep learning. *IEEE Transactions on Industrial Electronics*, 68(4):3577–3587, 2020.

[134] Masoud S Bahraini, Ahmad B Rad, and Mohammad Bozorg. Slam in dynamic environments: A deep learning approach for moving object tracking using ml-ransac algorithm. *Sensors*, 19(17):3699, 2019.

[135] Guangqiang Li, Lei Yu, and Shumin Fei. A deep-learning real-time visual slam system based on multi-task feature extraction network and self-supervised feature points. *Measurement*, 168:108403, 2021.

[136] Lu Sheng, Dan Xu, Wanli Ouyang, and Xiaogang Wang. Unsupervised collaborative learning of keyframe detection and visual odometry towards monocular deep slam. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4302–4311, 2019.

[137] Joao Carlos Virgolino Soares, Marcelo Gattass, and Marco Antonio Meggiolaro. Visual slam in human populated environments: exploring the trade-off between

accuracy and speed of yolo and mask r-cnn. In *2019 19th International Conference on Advanced Robotics (ICAR)*, pages 135–140. IEEE, 2019.

[138] Dan Xu, Andrea Vedaldi, and Joao F Henriques. Moving slam: Fully unsupervised deep learning in non-rigid scenes. *arXiv preprint arXiv:2105.02195*, 2021.

[139] Hudson Martins Silva Bruno and Esther Luna Colombini. Lift-slam: a deep-learning feature-based monocular visual slam method. *Neurocomputing*, 455:97–110, 2021.

[140] Saba Arshad and Gon-Woo Kim. Role of deep learning in loop closure detection for visual and lidar slam: A survey. *Sensors*, 21(4):1243, 2021.

[141] Rohit Mittal, Vibhakar Pathak, Geeta Chhabra Gandhi, Amit Mithal, and Kamlesh Lakhwani. Application of machine learning in slam algorithms. In *Machine Learning for Sustainable Development*, pages 147–160. De Gruyter, 2021.

[142] Peter Karkus, Shaojun Cai, and David Hsu. Differentiable slam-net: Learning particle slam for visual navigation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2815–2825, 2021.

[143] Ruochen Wang, Ling Pei, Lei Chu, Qi Wu, Tao Li, Wenxian Yu, and Xujun Guan. Dvt-slam: Deep-learning based visible and thermal fusion slam. In *China Satellite Navigation Conference (CSNC 2021) Proceedings*, pages 394–403. Springer, 2021.

[144] Gang Li, Xiang Liao, Huilan Huang, Shaojian Song, Bin Liu, and Yawen Zeng. Robust stereo visual slam for dynamic environments with moving object. *IEEE Access*, 9:32310–32320, 2021.

[145] Ethan Rublee, Vincent Rabaud, Kurt Konolige, and Gary Bradski. Orb: An efficient alternative to sift or surf. In *2011 International conference on computer vision*, pages 2564–2571. Ieee, 2011.

[146] Giorgio Grisetti, Rainer Kümmerle, Hauke Strasdat, and Kurt Konolige. g2o: A general framework for (hyper) graph optimization. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), Shanghai, China*, pages 9–13, 2011.

[147] Bill Triggs, Philip Mclauchlan, Richard Hartley, and Andrew Fitzgibbon. Bundle Ajustment – A Modern Synthesis. In Bill Triggs, Andrew Zisserman, and Richard Szeliski, editors, *Vision Algorithms: Theory and Practice*, volume 1883 of *Lecture Notes in Computer Science (LNCS)*, pages 298–372. Springer-Verlag, 2000.

[148] David Nister and Henrik Stewenius. Scalable recognition with a vocabulary tree. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, volume 2, pages 2161–2168. Ieee, 2006.

[149] Raúl Mur-Artal and Juan D Tardós. Fast relocalisation and loop closing in keyframe-based slam. In *2014 IEEE International Conference on Robotics and Automation (ICRA)*, pages 846–853. IEEE, 2014.

[150] Wang Cheng, Oda Masahiro, Hayashi Yuichiro, Kitasaka Takayuki, Takabatake Hirotsugu, Mori Masaki, Honma Hirotoshi, Natori Hiroshi, and Mori Kensaku. Improvement of robustness of SLAM-based bronchoscope tracking by posture guided feature matching. In *International Journal of Computer Assisted Radiology and Surgery (CARS)*, volume 13, pages 11–12, 2018.

[151] Yukitaka Nimura. Pluto: a common platform for computer-aided diagnosis. *Med Imag Tech*, 26(3):187–191, 2008.

[152] Zhengyou Zhang. A flexible new technique for camera calibration. *IEEE Transactions on pattern analysis and machine intelligence*, 22(11):1330–1334, 2000.

[153] Marc Levoy. Display of surfaces from volume data. *IEEE Computer graphics and Applications*, 8(3):29–37, 1988.

[154] Carles Sánchez, Jorge Bernal, Debora Gil, and F Javier Sánchez. On-line lumen centre detection in gastrointestinal and respiratory endoscopy. In *Workshop on Clinical Image-Based Procedures*, pages 31–38. Springer, 2013.

[155] Debora Gil, Antonio Esteban-lansaque, Agnés Borràs, Esmitt Ramírez, and Carles Sánchez. Intraoperative extraction of airways anatomy in videobronchoscopy. *IEEE Access*, 2020.

[156] Oliver Tonet, TU Ramesh, Giuseppe Megali, and Paolo Dario. Tracking endoscopic instruments without localizer: image analysis-based approach. *Studies in health technology and informatics*, 119:544–549, 2006.

[157] Xenophon Zabulis, Antonis A Argyros, and Dimitris P Tsakiris. Lumen detection for capsule endoscopy. In *2008 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 3921–3926. IEEE, 2008.

[158] K Vijayan Asari. A fast and accurate segmentation technique for the extraction of gastrointestinal lumen from endoscopic images. *Medical engineering & physics*, 22(2):89–96, 2000.

[159] S. Morimitsu, M. Oda, K. Tanaka, H. Takabatake, M. Mori, H. Natori, and K. Mori. Blood vessel segmentation from laparoscopic video using ConvLSTM U-Net. *Int J CARS*, 15(Suppl 1):S63–S64, 2020.

[160] Daniil Pakhomov, Vittal Premachandran, Max Allan, Mahdi Azizian, and Nassir Navab. Deep residual learning for instrument segmentation in robotic surgery. In *International Workshop on Machine Learning in Medical Imaging*, pages 566–573. Springer, 2019.

[161] Diego Bravo, Josue Ruano, Martín Gómez, and Eduardo Romero. Automatic polyp detection and localization during colonoscopy using convolutional neural networks. In *Medical Imaging 2020: Computer-Aided Diagnosis*, volume 11314, page 113143A. International Society for Optics and Photonics, 2020.

[162] J. Zhu, T. Park, P. Isola, and A. A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *2017 ICCV*, pages 2242–2251, 2017.

[163] M. Oda, K. Tanaka, H. Takabatake, M. Mori, H. Natori, and K. Mori. Realistic endoscopic image generation method using virtual-to-real image-domain translation. *Healthc. Technol. Lett.*, 6(6):214–219, 2019.

[164] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In Nassir Navab, Joachim Hornegger, William M. Wells, and Alejandro F. Frangi, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pages 234–241, Cham, 2015. Springer International Publishing.

[165] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Instance normalization: The missing ingredient for fast stylization. *arXiv preprint arXiv:1607.08022*, 2016.

[166] N. Otsu. A threshold selection method from gray-level histograms. *IEEE T SYST MAN CY-S*, 9(1):62–66, 1979.

[167] Cheng Wang, Masahiro Oda, Yuichiro Hayashi, Benjamin Villard, Takayuki Kitasaka, Hirotsugu Takabatake, Masaki Mori, Hirotoshi Honma, Hiroshi Natori, and Kensaku Mori. A visual slam-based bronchoscope tracking scheme for bron-

choscopic navigation. *International Journal of Computer Assisted Radiology and Surgery*, 15(10):1619–1630, 2020.

[168] Riyoko Shinohara, Kensaku Mori, Daisuke Deguchi, Takayuki Kitasaka, Yasuhito Suenaga, Hirotsugu Takabatake, Masaki Mori, and Hiroshi Natori. Branch identification method for ct-guided bronchoscopy based on eigenspace image matching between real and virtual bronchoscopic images. In *Medical Imaging 2006: Physiology, Function, and Structure from Medical Images*, volume 6143, page 614314. International Society for Optics and Photonics, 2006.

[169] Antonio Esteban-Lansaque, Carles Sánchez, Agnés Borràs, Marta Diez-Ferrer, Antoni Rosell, and Debora Gil. Stable anatomical structure tracking for video-bronchoscopy navigation. In *Workshop on Clinical Image-Based Procedures*, pages 18–26. Springer, 2016.

[170] Abraham Savitzky and Marcel JE Golay. Smoothing and differentiation of data by simplified least squares procedures. *ANAL. CHEM.*, 36(8):1627–1639, 1964.

[171] Richard Hartley and Andrew Zisserman. *Multiple view geometry in computer vision*. Cambridge university press, 2003.

Chapter 6

Publications

Journal papers

(1) **Cheng Wang**, Masahiro Oda, Yuichiro Hayashi, Benjamin Villard, Takayuki Kitasaka, Hirotsugu Takabatake, Masaki Mori, Hirotoshi Honma, Hiroshi Natori, Kensaku Mori. "A visual SLAM-based bronchoscope tracking scheme for bronchoscopic navigation." International Journal of Computer Assisted Radiology and Surgery. 15, pp. 1619–1630. (IF: 2.4) 2020.

(2) **Cheng Wang**, Masahiro Oda, Yuichiro Hayashi, Takayuki Kitasaka, Hirotsugu Takabatake, Masaki Mori, Hirotoshi Honma, Hiroshi Natori, Kensaku Mori. "Depth-based branching level estimation for bronchoscopic navigation." (Accepted by International Journal of Computer Assisted Radiology and Surgery)

Conference papers (International)

(1) **Cheng Wang**, Masahiro Oda, Yuichiro Hayashi, Takayuki Kitasaka, Hirotsugu Takabatake, Masaki Mori, Hirotoshi Honma, Hiroshi Natori, Kensaku Mori. "Depth-

based branching level estimation for bronchoscopic navigation." International Journal of Computer Assisted Radiology and Surgery (CARS 2021). 2021.

(2) **Cheng Wang**, Masahiro Oda, Yuichiro Hayashi, Takayuki Kitasaka, Hirotsugu Takabatake, Masaki Mori, Hirotoshi Honma, Hiroshi Natori, Kensaku Mori. "Bronchial orifice segmentation on bronchoscopic video frames based on generative adversarial depth estimation." Medical Imaging 2021: Image-Guided Procedures, Robotic Interventions, and Modeling. Vol. 115980N. 2021.

(3) **Cheng Wang**, Masahiro Oda, Yuichiro Hayashi, Takayuki Kitasaka, Hirotsugu Takabatake, Masaki Mori, Hirotoshi Honma, Hiroshi Natori, Kensaku Mori. "Improved visual SLAM for bronchoscope tracking and registration with pre-operative CT images." Medical Imaging 2020: Image-Guided Procedures, Robotic Interventions, and Modeling. Vol. 11315, pp. 1131518. 2020

(4) **Cheng Wang**, Masahiro Oda, Yuichiro Hayashi, Takayuki Kitasaka, Hirotsugu Takabatake, Masaki Mori, Hirotoshi Honma, Hiroshi Natori, Kensaku Mori. "TACR: trajectory-airways centerline registration and its application to bronchoscopic navigation." Asian Conference on Computer Aided Surgery (ACCAS 2019). 2019.

(5) **Cheng Wang**, Masahiro Oda, Yuichiro Hayashi, Takayuki Kitasaka, Hirotsugu Takabatake, Masaki Mori, Hirotoshi Honma, Hiroshi Natori, Kensaku Mori. "Evaluation on reconstruction accuracy of visual SLAM based bronchoscope tracking." International Journal of Computer Assisted Radiology and Surgery (CARS 2019). Vol. 14, Sup. 1, pp. S24-25. 2019.

(6) **Cheng Wang**, Masahiro Oda, Yuichiro Hayashi, Takayuki Kitasaka, Hirotsugu Takabatake, Masaki Mori, Hirotoshi Honma, Hiroshi Natori, Kensaku Mori. "Visual SLAM for bronchoscope tracking and bronchus reconstruction in bronchoscopic

navigation." Medical Imaging 2019: Image-Guided Procedures, Robotic Interventions, and Modeling. Vol.10951, pp. 108510A. 2019.

(7) **Cheng Wang**, Masahiro Oda, Yuichiro Hayashi, Takayuki Kitasaka, Hirotsugu Takabatake, Masaki Mori, Hirotoshi Honma, Hiroshi Natori, Kensaku Mori. "Improvement of robustness of SLAM-based bronchoscope tracking by posture guided feature matching." International Journal of Computer Assisted Radiology and Surgery (CARS 2018). Vol. 13, Sup. 1, pp. s11-12. 2018.

(8) **Cheng Wang**, Masahiro Oda, Yuichiro Hayashi, Holger Reinhard Roth, Kazunari Misawa, Kensaku Mori. Motion vector for outlier elimination in feature matching and its application in SLAM based laparoscopic tracking. International conference on Medical Image Computing and computer Assisted Intervention (MICCAI 2017). Computer Assisted and Robotic Endoscopy and Clinical Image-Based Procedures, LNCS 10550, pp.60-69. 2017.

(9) Shuai Jiang, **Cheng Wang**, Yuichiro Hayashi, Masahiro Oda, Takayuki Kitasaka, Kazunari Misawa, Kensaku Mori, Automated hand eye calibration in laparoscope holding robot for robot assisted surgery," IFMIA 2019, Proceedings of SPIE 11049, pp.1104949-1-5. 2019.

Conference papers (Japan)

(1) **Cheng Wang**, Masahiro Oda, Yuichiro Hayashi, Takayuki Kitasaka, Hirotsugu Takabatake, Masaki Mori, Hirotoshi Honma, Hiroshi Natori, Kensaku Mori. "Evaluation of generative adversarial depth estimation for bronchoscopic navigation," Proceedings of Japan Society of Computer Aided Surgery (JSCAS 2020), 22(4), 20:38-339,2020.

(2) **Cheng Wang**, Masahiro Oda, Yuichiro Hayashi, Takayuki Kitasaka, Hirotsugu Takabatake, Masaki Mori, Hirotoshi Honma, Hiroshi Natori, Kensaku Mori. "Preliminary Study on Application of ORB-SLAM in Bronchoscope Tracking," Proceedings of Japan Society of Computer Aided Surgery (JSCAS 2018), 20(4):342-343, 2018.

(3) **Cheng Wang**, Masahiro Oda, Yuichiro Hayashi, Takayuki Kitasaka, Hirotsugu Takabatake, Masaki Mori, Hirotoshi Honma, Hiroshi Natori, Kensaku Mori. "Evaluation of Adjacent Reconstruction in SLAM-based Visual Tracking." Proceedings of Japanese Society of Medical Imaging Technology (JAMIT 2017), OP13-6,2018.

(4) **Cheng Wang**, Masahiro Oda, Yuichiro Hayashi, Kensaku Mori. "Improvement on Robustness of ORB-SLAM Based Surgical Navigation System by Building Submap." Proceedings of Japanese Society of Medical Imaging Technology (JAMIT 2017), OP2-6,2017.

(5) **Cheng Wang**, Masahiro Oda, Yuichiro Hayashi, Kensaku Mori. "Preliminary Study on Application of ORB-SLAM in Bronchoscope Tracking." Proceedings of Japan Society of Computer Aided Surgery (JSCAS 2017), 19(4):324-325,2017.

(6) **Cheng Wang**, Masahiro Oda, Yuichiro Hayashi, Takayuki Kitasaka, Hirotsugu Takabatake, Masaki Mori, Hirotoshi Honma, Hiroshi Natori, Kensaku Mori. "Investigation on the condition of using adjacent reconstruction in visual bronchoscope tracking."the Institute of Electronics, Information and Communication Engineers (IEICE), 118(286):27-32, 2018.

(7) **Cheng Wang**, Masahiro Oda, Yuichiro Hayashi, Hirotsugu Takabatake, Masaki Mori, Hirotoshi Honma, Hiroshi Natori, Kensaku Mori. "Study on the Robustness of ORB-SLAM Based Outlier Elimination in Bronchoscope Tracking – RANSAC +

EPnP for Outlier Detection." the Institute of Electronics, Information and Communication Engineers (IEICE), 117(281):47-52,2017.

(8) **Cheng Wang**,Mohammad Eshghi, Masahiro Oda, Yuichiro Hayashi,Kazunari Misawa, Zhengang Jiang, Kensaku Mori. "Preliminary Study on Application of ORB-SLAM in Laparoscope Navigation System." Proceedings of Japan Society of Computer Aided Surgery (JSCAS 2016), 18(4):355-356, 2016.