

Development and Comparison of New Hybrid Motion Tracking for Bronchoscopic Navigation

Xióngbiāo Luó^a, Marco Feuerstein^{a,b}, Daisuke Deguchi^a,
Takayuki Kitasaka^{c,d}, Hirotsugu Takabatake^e, Kensaku Mori^{a,d,f}

^aGraduate School of Information Science, Nagoya University, Japan;

^bComputer Aided Medical Procedures (CAMP), Technische Universität München, Germany;

^cFaculty of Information Science, Aichi Institute of Technology, Japan;

^dMEXT Innovation Center for Preventive Medical Engineering, Nagoya University, Japan;

^eSapporo-Minami-Sanjo Hospital, Japan;

^fStrategy Office, Information and Communications Headquarters, Nagoya University, Japan

Abstract

This paper presents a new hybrid camera motion tracking method for bronchoscopic navigation combining SIFT, epipolar geometry analysis, Kalman filtering, and image registration. In a thorough evaluation, we compare it to state-of-the-art tracking methods. Our hybrid algorithm for predicting bronchoscope motion uses SIFT features and epipolar constraints to obtain an estimate for inter-frame pose displacements and Kalman filtering to find an estimate for the magnitude of the motion. We then execute bronchoscope tracking by performing image registration initialized by these estimates. This procedure registers the actual bronchoscopic video and the virtual camera images generated from 3D chest CT data taken prior to bronchoscopic examination for continuous bronchoscopic navigation. A comparative assessment of our new method and the state-of-the-art methods is performed on actual patient data and phantom data. Experimental results from both datasets demonstrate a significant performance boost of navigation using our new method. Our hybrid method is a promising means for bronchoscope tracking, and outperforms other methods based solely on Kalman filtering or image features and image registration.

Keywords:

Bronchoscopic navigation system, Camera motion estimation, Virtual bronchoscopy, Image registration, Bronchoscope tracking

1. Introduction

In minimally invasive diagnosis and treatment of lung and bronchus cancer, physicians must obtain useful information for the assessment of suspicious tumors and site localization. The most popular means are 3D multidetector computed-tomography (CT) data showing the high resolution images of the airways and bronchoscopy providing endoluminal images in real time. During bronchoscopic examinations, physicians need to see previously acquired CT data aligned with the bronchoscopic video. Unfortunately, it is very difficult for bronchoscopists to manually fuse that information because of the complexity of the bronchial tree and that no direct connection exists between the CT images and real bronchoscopic images. Therefore, we must develop a bronchoscopic guidance system to assist physicians during this mental burden. Such a bronchoscopic navigation system needs to accurately track the position and orientation of the bronchoscope camera inside the patient's airway tree relative to the CT and in real time.

For bronchoscope tracking, two main approaches (or their combination) have been proposed in the literature up to now: (a) electromagnetic sensor-based and (b) image registration-based tracking. Sensor-based guidance utilizes a sensing coil (sensor) attached to the tip of the bronchoscope and localized by an electromagnetic tracking system, such as the commer-

cially available superDimension navigation system (Solomon et al., 2000; Gildea et al., 2006; Schwarz et al., 2006; Becker et al., 2005). However, such navigation systems suffer from the following bottlenecks: (a) sensitivity to localization problems resulting from any patient movement (i.e., respiratory motion or coughing). (b) measurement inaccuracies because of magnetic field distortion caused by ferrous metals or conductive material within or close to the working volume; (c) difficulties when integrating sensors into ultra-thin bronchoscopes (i.e., 2.8 mm in external diameter and 1.2 mm in working channel diameter) because of space restrictions.

Image registration-based tracking for bronchoscopic navigation (Peters and Cleary, 2008), which is a very active topic of research is also the topic of our paper. It is often comprised of two major stages: camera motion estimation and image registration. In the first stage, we distinguish between two kinds of methods: filtering-based and feature-based motion estimation. Since bronchoscope tracking is a nonlinear filtering problem that requires nonlinear filtering approaches for its solution, methods based on Bayesian or motion filtering have already been proposed. Nagao et al. (2004) used a Kalman-based estimator to increase the speed and accuracy of their registration algorithm. Deligianni et al. (2005) employed the CONDensation algorithm to avoid trapping the registration in local minima. Al-

though filtering-based methods increase the general accuracy and robustness of registration involved in bronchoscope tracking, they usually do not estimate the rotational part of the bronchoscope motion. However, feature-based methods for bronchoscope motion prediction seem promising for bronchoscope tracking. In [Mori et al. \(2002\)](#) and [Rai et al. \(2008\)](#), optical flow patterns compute bronchoscope motion between consecutive real bronchoscopic (RB) images as a pre-registration step for bronchoscopic navigation. Although good performance was shown, it remains difficult to get an accurate estimate for the insertion depth of the bronchoscope, or, in other words, for the magnitude of translation between successive frames.

To improve the results of the first stage, image registration can be invoked to precisely estimate the position and the orientation of the bronchoscope. Therefore, a virtual bronchoscopic (VB) view generated by placing a virtual camera inside the 3D CT data that resembles the current RB image is searched for by altering the rendering parameters used to generate this virtual image. During image registration, the image similarity between VB and RB images is calculated. Numerous papers have been published on this topic. [Bricault et al. \(1998\)](#) presented their pioneering work in registering RB and VB images to assist transbronchial biopsies and utilized bifurcations inspected in the bronchial tree of the patient to calculate the camera position corresponding to RB images. However, since their method is constrained to bifurcations, it cannot tackle cases where this information cannot be clearly observed. [Helferty and Higgins \(2001\)](#) proposed a method for bronchoscope tracking using image registration based on normalized mutual information and greatly improved the tracking speed and accuracy by employing 2D/2D matching frameworks to obtain 3D camera motion ([Helferty et al., 2007](#)). [Deligianni et al. \(2006\)](#) proposed bronchoscope tracking using a position sensor and a pq -based registration technique and improved its accuracy and stability by modeling respiratory motion with an active shape model. Their method generates VB images that most resemble RB images by recovering a bidirectional reflectance distribution function (BRDF) from registered RB and VB images ([Chung et al., 2006](#)). Our group developed an image registration-based approach to continuously track the camera motion by spatially aligning RB and VB images (see [Figure 1](#)) on the basis of the selective image similarity measure $MoMSE$ ([Deguchi et al., 2009](#)). Although our previously proposed method was accurate and effective for some test cases, it still suffers from such problems as an inability to recover from registration errors and abrupt (jumpy) transitions between successive frames.

This paper explores a hybrid method for predicting bronchoscope motion and first uses features obtained by scale invariant feature transform (SIFT) ([Lowe, 2004](#)) to estimate the inter-frame pose displacements (up to scale) based on epipolar constraints and Kalman filtering to get position estimates and then performs intensity-based image registration initialized by these estimates for improving the performance of bronchoscopic guidance.

Our contributions can be summarized as follows. First, we introduce SIFT-based motion estimation, which employs SIFT features to predict camera motion in bronchoscopic navigation.

Combining it with a Kalman filtering-based method, we construct a new hybrid method for bronchoscope motion estimation that can accomplish better initialization for a subsequent image registration process than other methods and hence improve accuracy. Second, a new method for scale factor estimation (i.e., estimation of the magnitude of the translation between two successive frames) is described. We demonstrate that the tracking results of feature-based methods are sensitive to the scale factor of the translation vector when performing 2D/3D pose prediction. In this paper for the first time we predict the scale factor by utilizing Kalman filtering. Third, we present a thorough assessment of different methods of bronchoscope motion estimation. Initially we compare two kinds of motion estimation methods: (a) filtering-based and (b) feature-based. Based on this comparison, we formulate a better way to predict bronchoscope motion to improve the robustness and accuracy of pose estimation. Additionally, we provide a complete procedure to acquire bronchoscopic ground truth data by combining camera and hand-eye calibrations and magnetic tracking.

The remainder of this paper is organized as follows. The parameters required to describe the camera motion are defined in [Section 2](#). Our new hybrid bronchoscope tracking method is described in [Section 3](#). In [Section 3.1](#), we recall the paradigm of camera motion estimation by feature-based methods. A new scale factor determination method based on Kalman filtering is presented in [Section 3.2](#). [Section 3.3](#) briefly summarizes the parameters used to generate VB images and to register them to their real counterparts. [Section 4](#) details the experimental environment and assessment scheme. Finally, [Section 5](#) summarizes and [Section 6](#) discusses the assessment results, before [Section 7](#) concludes this work.

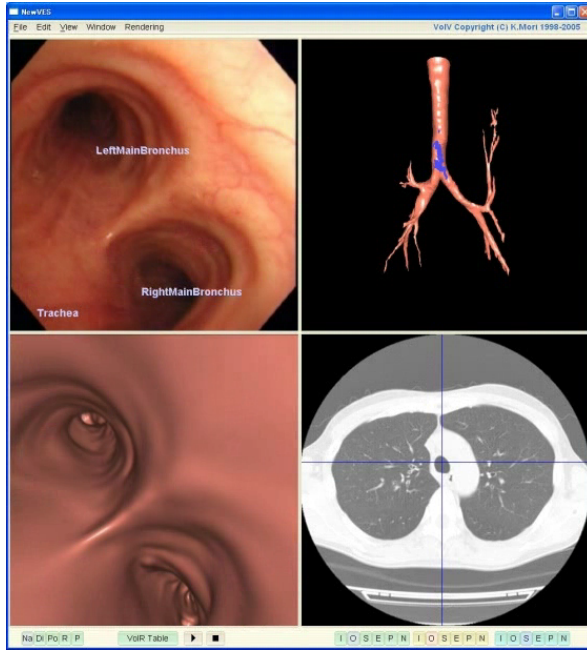
2. Definition

While the input of a bronchoscopic navigation system usually is a set of CT slices and bronchoscopic video from the same patient, its output is a sequence of VB renderings, where the estimated bronchoscope position and orientation are represented in CT coordinates. Bronchoscope tracking can be considered a search procedure that determines a time series of VB images corresponding to the RB images with maximum similarity between them.

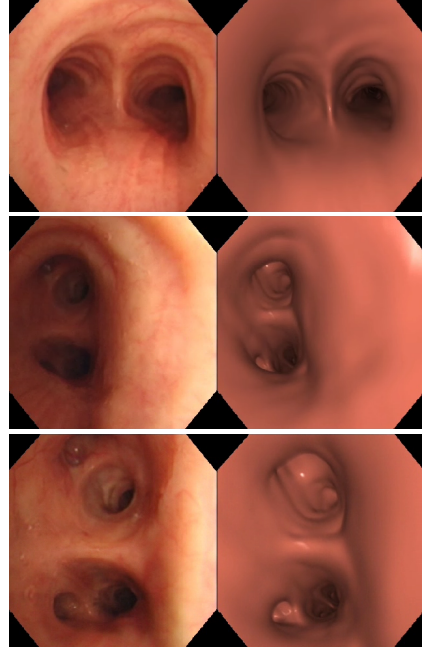
We denote the extrinsic camera parameters for the (i) -th rendered VB image as $\mathbf{Q}^{(i)}$ that correspond to the RB camera position and orientation and are described by

$$\mathbf{Q}^{(i)} = \begin{pmatrix} \mathbf{R}^{(i)} & \mathbf{t}^{(i)} \\ \mathbf{0}^T & 1 \end{pmatrix}, \quad (1)$$

where $\mathbf{R}^{(i)}$ and $\mathbf{t}^{(i)}$ represent a rotation matrix and a translation vector of the bronchoscopic camera in CT coordinates, respectively. This homogeneous matrix depicts the transformation from the camera coordinate system to the CT coordinate system. When the bronchoscope performs small inter-frame movements comprising translational motion $\Delta\mathbf{t}^{(i)}$ and rotational motion $\Delta\mathbf{R}^{(i)}$ between the $(i-1)$ -th and (i) -th RB images, the up-



(a)



(b)

Figure 1: Composite view of our bronchoscopic navigation system and fusion between RB and VB images. (a) Our bronchoscopic navigation system contains four windows for displaying useful information. RB video (upper left), a VB image (lower left) generated from previously acquired CT data and corresponding to the RB image, a bird's view (upper right) displaying the bronchoscope trajectory, and a slice view (lower right) showing an axial slice, in which a cross cursor marks the current bronchoscope location. All windows were synchronized to provide critical visual and structural information during bronchoscopy. (b) Other examples of corresponding RB (left column) and VB images (right column). VB images on right are rendered at estimated camera positions of bronchoscope. These images were rendered in the same way as those in the lower-left view of (a).

dated camera motion parameters are formulated as

$$\mathbf{Q}^{(i)} = \mathbf{Q}^{(i-1)} \Delta \mathbf{Q}^{(i)} = \mathbf{Q}^{(i-1)} \begin{pmatrix} \Delta \mathbf{R}^{(i)} & \Delta \mathbf{t}^{(i)} \\ \mathbf{0}^T & 1 \end{pmatrix}, \quad (2)$$

where $\Delta \mathbf{t}^{(i)} = (\Delta t_x^{(i)}, \Delta t_y^{(i)}, \Delta t_z^{(i)})^T$ defines a translation vector. Relative rotation matrix $\Delta \mathbf{R}^{(i)}$ is constructed by Euler rotation angles θ , ϕ , and ψ of the camera around the x -, y -, and z -axes and is described by

$$\Delta \mathbf{R}^{(i)} = \begin{pmatrix} C_3 C_2 & C_3 S_2 S_1 - S_3 C_1 & C_3 S_2 C_1 + S_3 S_1 \\ S_3 C_2 & S_3 S_2 S_1 + C_3 C_1 & S_3 S_2 C_1 - C_3 S_1 \\ -S_2 & C_2 S_1 & C_2 C_1 \end{pmatrix} \quad (3)$$

where the variables of this matrix are defined as: $S_1 = \sin \theta$, $S_2 = \sin \phi$, $S_3 = \sin \psi$, $C_1 = \cos \theta$, $C_2 = \cos \phi$, and $C_3 = \cos \psi$.

Figure 2 further shows the relationship between the camera and the CT coordinates. Inter-frame movement $\Delta \mathbf{Q}^{(i)}$ is represented in the camera coordinate system at the $(i-1)$ -th frame and indicates the change of camera position and orientation from the $(i-1)$ -th to i -th frames. RB camera motion tracking is obtained by continuously estimating $\Delta \mathbf{t}^{(i)}$ and $\Delta \mathbf{R}^{(i)}$ under the assumption that $\mathbf{Q}^{(0)}$ is given.

Additionally, we need the intrinsic camera parameters for the generation of VB images by volume rendering and for epipolar

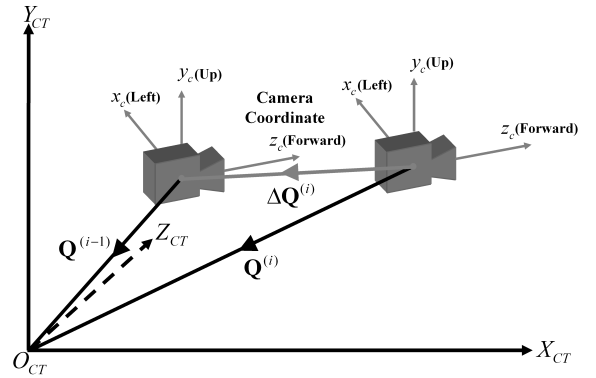


Figure 2: Relationship of coordinate systems and transformations during predictive camera motion tracking

geometry analysis. We define intrinsic matrix \mathbf{K} as

$$\mathbf{K} = \begin{pmatrix} f_x & 0 & p_x \\ 0 & f_y & p_y \\ 0 & 0 & 1 \end{pmatrix}, \quad (4)$$

where f_x and f_y are the focal lengths in units of horizontal and vertical pixels and $(p_x, p_y)^T$ is the principal point. We also model the tangential and radial distortion of the camera lens according to (Heikkilä and Silvén, 1997).

3. Hybrid Bronchoscope Tracking

In this section, we present all the components of our hybrid approach to bronchoscope tracking. In this context, *hybrid* refers to a combination of feature-based tracking, epipolar geometry analysis, and filter-based motion tracking for constructing a new motion estimation algorithm and its integration into image registration for bronchoscope tracking¹. The new hybrid motion tracking method combines filtering- and feature-based approaches by fusing their respective advantages to improve the overall performance of bronchoscope tracking. Figure 3 shows a flowchart of our hybrid tracking method. We first extract stable scale invariant feature transform (SIFT) features from the current bronchoscopic video image and find matching features in the previous image. The matches are used for epipolar geometry analysis, which estimates the relative camera motion up to scale. To recover this unknown scale, we additionally apply Kalman filtering, which gives an estimation of the relative camera translation. By fusing the epipolar analysis and the Kalman filtering estimates to a full 6 degrees of freedom (DOF) camera motion matrix and multiplying it onto the camera pose matrix obtained for the previous image, we get an estimate for the current camera pose to initialize the final image registration step.

Note that our hybrid bronchoscope tracking approach exhibits similarities to simultaneous localization and mapping (SLAM) (Durrant-Whyte and Balley, 2006), particularly MonoSLAM (Davison et al., 2007), a technique developed to permit a robot to build a map within an unknown environment or update a map within a known environment using a single camera while simultaneously determining its 3D motion trajectory within this map. Our method also recovers a 3D trajectory inside the CT images when the bronchoscope is moved along the patient’s airways, and, similar to MonoSLAM, utilizes image features as well as Kalman filtering for that purpose. However, here, we cannot rely only on image features and Kalman filtering, since we additionally need to deal with respiratory motion and camera trajectory mainly along the viewing direction of the camera, which has not been addressed in previous work using MonoSLAM in endoscopy (Grasa et al., 2009). To successfully address these issues, our method fuses the essential information from feature detection, epipolar analysis, and Kalman filtering with image registration to achieve more stable camera tracking.

3.1. Feature-based Motion Estimation

Feature-based algorithms for structure and motion computation, which are now widely developed for tracking a small number of salient features in long image sequences, can be divided into two major steps: (a) feature extraction and (b) feature tracking. The first stage detects features from camera images and identifies corresponding features in each image of the sequence that can be utilized in the second stage to calculate the relative camera motion (up to scale). In this paper, we focus on deriving features from bronchoscopic video sequences and utilize a SIFT-based method for feature tracking in RB images.

¹This is contrary to our previous hybrid method that combined electromagnetic tracking and image registration (Mori et al., 2005).

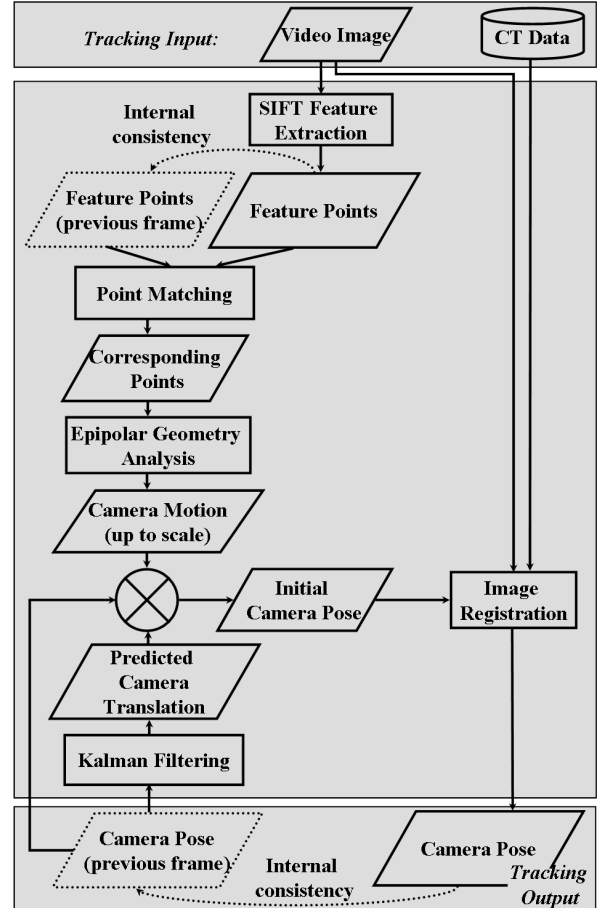


Figure 3: Processing flowchart of our hybrid motion tracking method

3.1.1. Feature Extraction

For feature-based motion estimation, the first important step is the extraction of features. Currently one of the best-known methods for extracting image features is the SIFT algorithm. SIFT returns distinctive features from keypoints that are invariant to image scale and rotation (Lowe, 2004). We introduce SIFT in our hybrid motion estimator for predicting bronchoscope motion. It uses SIFT features and epipolar geometric constraints to calculate the relative translation up to scale and the rotation between the previous and current poses of the bronchoscope camera. Given RB image sequences observing the bronchial structure from different viewpoints, the points of interest are detected on the difference of the Gaussian pyramid, and each point is represented by a SIFT descriptor composed of a 128-dimensional vector containing local orientation histograms around the keypoint’s position.

To speed up the feature tracking procedure in an RB image sequence, we always store the feature context of a processed RB image to recall it during the next iteration.

3.1.2. Point Matching Strategy

Given two consecutive RB images, we extract two sets of interest points and calculate the point descriptors. Before we can use these interest points for bronchoscope tracking, we must

identify the corresponding points between successive RB images.

There are usually three strategies for performing point matching, as shown in Mikolajczyk and Schmid (2005). The simplest choice is threshold-based matching: two points are considered to match if the distance between their descriptors is equal or lower than a threshold. In this case, one descriptor in the previous frame may correspond to several descriptors in the current frame. Another choice is to add a nearest neighbor constraint to obtain one-to-one correspondences. To achieve even more effective matching, a third strategy can be utilized that computes the ratio between the distance of the nearest and the second-nearest neighbors and thresholds this ratio to discard false matches. The third strategy is described in Lowe (2004).

In our experiments, we performed point matching based on the third matching strategy by computing the Euclidean distance between the descriptors and the above distance ratio to obtain L matches. Based on the experiments performed in Lowe (2004), a distance ratio threshold between about 0.2 and 0.8 gives good results. We set the threshold to 0.6, which was the default for the demo implementation provided by Lowe², since this provided a good compromise between the elimination of enough false matches and not discarding too many correct matches.

Even though the third matching strategy is effective, obtaining unsuitable point pairs is still unavoidable. Therefore, we implement a simple outlier detection mechanism based on the standard deviation of the distances between corresponding points.

Given two sets of matching points $\{\mathbf{p}_l^{(i-1)} = (u_l^{(i-1)}, v_l^{(i-1)})^T\}_{l=1}^L$ and $\{\mathbf{p}_l^{(i)} = (u_l^{(i)}, v_l^{(i)})^T\}_{l=1}^L$, which are detected in two different RB images based on the third matching strategy, respectively, where u_l and v_l are the pixel coordinates in RB images, we calculate Euclidean distances d_l of all point pairs $\mathbf{p}_l^{(i-1)}$ and $\mathbf{p}_l^{(i)}$

$$d_l = \|\mathbf{p}_l^{(i-1)} - \mathbf{p}_l^{(i)}\| = \sqrt{(u_l^{(i-1)} - u_l^{(i)})^2 + (v_l^{(i-1)} - v_l^{(i)})^2}. \quad (5)$$

Then we calculate mean \bar{d} and standard deviation σ of distance set $\{d_l, l = 1, \dots, L\}$. Now we define a non-zero threshold as $(\bar{d} + \sigma)$ to determine whether the current point pair is correct. Pair $(\mathbf{p}_l^{(i-1)}, \mathbf{p}_l^{(i)})$ is assumed to be correct only if

$$d_l \leq (\bar{d} + \sigma), \quad (6)$$

otherwise, it will be rejected. Applying our outlier rejection, we reduce the number of matches from L to K . $(\bar{d} - \sigma)$ is not considered a lower threshold for determining true or false matchings since standard deviation σ is greater than mean \bar{d} in successive frames with small movement between them. Figure 4 shows an example of the matching result between two consecutive frames.

3.1.3. Epipolar Geometry Analysis

Inter-frame movement $\Delta\mathbf{Q}^{(i)}$ is determined based on epipolar constraints that our corresponding point pairs $(\mathbf{p}_k^{(i-1)}, \mathbf{p}_k^{(i)})$

must satisfy ($k = 1, \dots, K$: K is the total number of matching points found in Section 3.1.2). The epipolar geometry, which is the intrinsic projective geometry between two views, can be algebraically represented by fundamental matrix \mathbf{F} (Hartley and Zisserman, 2004). For any pair of corresponding points $\mathbf{p}_k^{(i-1)} \leftrightarrow \mathbf{p}_k^{(i)}$ in two images, fundamental matrix \mathbf{F} can be delineated by

$$\mathbf{p}_k^{(i)T} \mathbf{F} \mathbf{p}_k^{(i-1)} = 0. \quad (7)$$

Given enough (at least 7) point matches $\mathbf{p}_k^{(i-1)} \leftrightarrow \mathbf{p}_k^{(i)}$, Eq. 7 can compute unknown matrix \mathbf{F} , for example, by the normalized eight-point algorithm of Hartley (Hartley and Zisserman, 2004).

Since we previously calibrated the bronchoscope to obtain its intrinsic camera parameters (stored in \mathbf{K}) (Zhang, 2000), we can now compute essential matrix $\mathbf{E} = \mathbf{K}^T \mathbf{F} \mathbf{K}$. We obtain inter-frame movement $\Delta\tilde{\mathbf{Q}}^{(i)}$ including translation unit vector $\Delta\tilde{\mathbf{t}}^{(i)}$ and rotation matrix $\Delta\tilde{\mathbf{R}}^{(i)}$ between the $(i-1)$ -th and (i) -th RB images by sequentially solving the following two equations (Mori et al., 2002):

$$\mathbf{E}^T \Delta\tilde{\mathbf{t}}^{(i)} = \mathbf{0}, \quad (8)$$

$$\Delta\tilde{\mathbf{R}}^{(i)} \mathbf{E}^T = [\Delta\tilde{\mathbf{t}}^{(i)}]_{\times}^T = \begin{bmatrix} 0 & -\Delta\tilde{t}_z^{(i)} & \Delta\tilde{t}_y^{(i)} \\ \Delta\tilde{t}_z^{(i)} & 0 & -\Delta\tilde{t}_x^{(i)} \\ -\Delta\tilde{t}_y^{(i)} & \Delta\tilde{t}_x^{(i)} & 0 \end{bmatrix}^T. \quad (9)$$

Note that essential matrix \mathbf{E} is determined only up to an arbitrary scale factor. Thus resulting translation vector $\Delta\tilde{\mathbf{t}}^{(i)}$ contains arbitrary scale factor $\tilde{\alpha}^{(i)} = |\Delta\tilde{\mathbf{t}}^{(i)}|$ that does not represent the real magnitude of translation $\alpha^{(i)} = |\Delta\mathbf{t}^{(i)}|$. It is difficult to estimate this scale factor for tracking the bronchoscope tip, since the main motion of the bronchoscope is along its viewing direction. This creates a very small baseline between the two camera positions of consecutive frames, making the depth and hence the scale factor estimation inaccurate even in the presence of metrics. In the next section, we present an effective method for determining this scale.

3.2. Kalman Filtering-Based Scale Factor Estimation

Filtering-based motion tracking methods in general estimate camera motion are not as accurate as feature-based methods because they only predict a new (unknown) state based on previous (known) states; feature-based methods can utilize known features in both the previous and current images. This in particular holds for the abrupt changes of the main motion direction. However, a prediction for only the magnitude of the motion (scale in this paper) can be more stable, since it is less dependent on the direction change itself.

Based on this idea, we construct our new hybrid motion estimator using SIFT features to obtain inter-frame pose displacements (with translation up to scale) and Kalman filtering to determine the unknown scale factors. In the following, we describe how to estimate the bronchoscope translation using Kalman filtering.

Kalman filtering, as one filtering-based estimator, can be described by a state-space model where observations occur and a

²The demo software is available from <http://www.cs.ubc.ca/~lowe/keypoints/>.

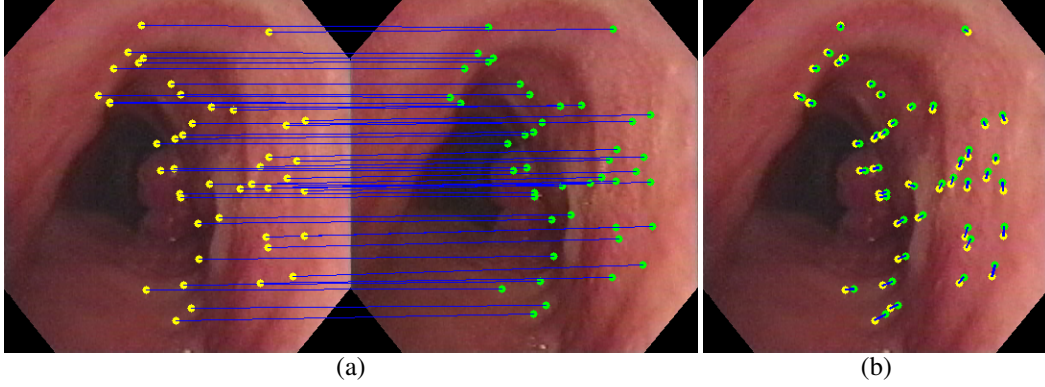


Figure 4: Example of point matching based on distance ratio between nearest and second-nearest neighbors and outlier detection. (a) shows yellow and green points that denote SIFT features at frame $(i-1)$ (left) and i (right), respectively. Blue lines connect corresponding points. (b) visualizes corresponding SIFT features (yellow and green points) at frame $(i-1)$ to see the motion of the bronchoscope camera from frame $(i-1)$ to i .

state is predicted at a discrete time. The state-space model may be formulated as

$$\mathbf{x}_i = \mathbf{B}_{i,i-1}\mathbf{x}_{i-1} + \mathbf{G}_{i-1}\mathbf{w}_{i-1}, \quad (10)$$

$$\mathbf{y}_{i-1} = \mathbf{H}_{i-1}\mathbf{x}_{i-1} + \mathbf{o}_{i-1}, \quad (11)$$

where $\mathbf{B}_{i,i-1}$ is a state transition matrix transferring unknown state vector \mathbf{x}_{i-1} to next vector \mathbf{x}_i from time $(i-1)$ to time i . During this estimation procedure, white noise \mathbf{w}_{i-1} and observation noise \mathbf{o}_{i-1} are assumed to be added. \mathbf{G}_{i-1} is the driving matrix. In Eq. 11, \mathbf{y}_{i-1} is designated as an observation signal and \mathbf{H}_{i-1} is the observation matrix.

The problem of Kalman filtering is to solve a set of mathematical equations for the unknown state vectors in an optimal manner that minimizes the estimated error covariance given part of the initial system state. Unknown state vector \mathbf{x}_i is predicted using observation \mathbf{y}_{i-1} to search for the minimal mean-square error estimation of state \mathbf{x}_{i-1} . The whole procedure implements predictor-corrector type estimation by solving the following formulations:

Corrector:

$$\hat{\mathbf{x}}_{i-1} = \hat{\mathbf{x}}_{i-2} + \hat{\mathbf{M}}_k(\mathbf{y}_{i-1} - \mathbf{H}_{i-1}\hat{\mathbf{x}}_{i-2}), \quad (12)$$

$$\hat{\mathbf{S}}_{i-1} = \hat{\mathbf{S}}_{i-2} - \hat{\mathbf{M}}_k\mathbf{H}_{i-1}\hat{\mathbf{S}}_{i-2}, \quad (13)$$

$$\hat{\mathbf{M}}_k = \hat{\mathbf{S}}_{i-2}\mathbf{H}_{i-1}^T(\mathbf{H}_{i-1}\hat{\mathbf{S}}_{i-2}\mathbf{H}_{i-1}^T + \mathbf{O}_{i-1})^{-1}, \quad (14)$$

Predictor:

$$\hat{\mathbf{x}}_i = \mathbf{B}_{i,i-1}\hat{\mathbf{x}}_{i-1}, \quad (15)$$

$$\hat{\mathbf{S}}_i = \mathbf{B}_{i,i-1}\hat{\mathbf{S}}_{i-1}\mathbf{B}_{i,i-1}^T + \mathbf{G}_{i-1}\mathbf{W}_{i-1}\mathbf{G}_{i-1}^T, \quad (16)$$

where \mathbf{O}_{i-1} and \mathbf{W}_{i-1} are the covariance matrices of \mathbf{o}_{i-1} and \mathbf{w}_{i-1} and $\hat{\mathbf{M}}_k$ is the Kalman gain. Eqs. 12 and 13 construct the corrector, estimated state $\hat{\mathbf{x}}_{i-1}$ and its error covariance matrix $\hat{\mathbf{S}}_{i-1}$ are calculated by updating previous estimation $\hat{\mathbf{x}}_{i-2}$ and error covariance matrix $\hat{\mathbf{S}}_{i-2}$. The predictor is built by Eqs. 15 and 16 that compute estimate $\hat{\mathbf{x}}_i$ and its error covariance matrix $\hat{\mathbf{S}}_i$ for approximating to state vector \mathbf{x}_i . More details on Kalman filter-based motion estimation are described in Kalman (1960), Welch (1996), and Welch and Bishop (2001).

In our studies, we utilize Kalman-based motion filtering to estimate the bronchoscope translation. Let $\mathbf{t}^{(i-1)}$ denote the translation vector of the bronchoscope at frame $(i-1)$. Assuming the bronchoscope is undergoing constant acceleration \mathbf{a} , we can define velocity $\mathbf{v}^{(i)}$ at frame i as:

$$\mathbf{v}^{(i)} = \frac{d\mathbf{t}}{dt} = \mathbf{v}^{(i-1)} + \mathbf{a}\Delta t, \quad (17)$$

where $d\mathbf{t}/dt$ is the derivative of translation $\Delta\mathbf{t} = \mathbf{t}^{(i)} - \mathbf{t}^{(i-1)}$ between frame $(i-1)$ and i and Δt is the time interval between the two frames. After integrating the derivative, we get

$$\Delta\mathbf{t} = \mathbf{v}^{(i-1)}\Delta t + \frac{1}{2}\mathbf{a}\Delta t^2, \quad (18)$$

and hence

$$\mathbf{t}^{(i)} = \mathbf{t}^{(i-1)} + \Delta\mathbf{t} = \mathbf{t}^{(i-1)} + \mathbf{v}^{(i-1)}\Delta t + \frac{1}{2}\mathbf{a}\Delta t^2. \quad (19)$$

In our bronchoscopic navigation system, $\mathbf{B}_{i,i-1}$, \mathbf{G}_{i-1} , and \mathbf{H}_{i-1} are assumed not to change over time because of a constant frame interval and the assumption of invariant acceleration in movements. We here define system state \mathbf{x}_i as $(\mathbf{t}^{(i)T}, \mathbf{v}^{(i)T}, \mathbf{a}^T)^T$. Based on Eqs. 17 and 19, we define

$$\mathbf{B}_{i,i-1} = \mathbf{B} = \begin{pmatrix} \mathbf{I}_{3\times 3} & (\Delta t)\mathbf{I}_{3\times 3} & (\frac{\Delta t^2}{2})\mathbf{I}_{3\times 3} \\ \mathbf{0}_{3\times 3} & \mathbf{I}_{3\times 3} & (\Delta t)\mathbf{I}_{3\times 3} \\ \mathbf{0}_{3\times 3} & \mathbf{0}_{3\times 3} & \mathbf{I}_{3\times 3} \end{pmatrix}, \quad (20)$$

$$\mathbf{G}_{i-1} = \mathbf{G} = (\mathbf{I}_{9\times 9}), \quad (21)$$

$$\mathbf{H}_{i-1} = \mathbf{H} = (\mathbf{I}_{3\times 3}, \mathbf{0}_{3\times 3}, \mathbf{0}_{3\times 3}), \quad (22)$$

$$\mathbf{w}_{i-1} = (\mathbf{w}_t^T, \mathbf{w}_v^T, \mathbf{w}_a^T)^T, \quad (23)$$

where $\mathbf{I}_{n\times n}$ and $\mathbf{0}_{n\times n}$ are the $n \times n$ identity and zero matrices. Assuming observation noise and its covariance matrix \mathbf{O}_{i-1} to be zero and the covariance of the white noise to be $\mathbf{W}_{i-1} = 0.05 \cdot \mathbf{I}_{9\times 9}$ (Nagao et al., 2004), we obtain the estimated system state by the state-space model and estimate the position of the bronchoscope:

$$\hat{\mathbf{t}}^{(i)} = \mathbf{H}\hat{\mathbf{x}}_i. \quad (24)$$

Next we use this estimation $\hat{\mathbf{t}}^{(i)}$ to compute the magnitude of the bronchoscope motion. Therefore, our Kalman-based scale factor determination can be expressed as

$$\hat{\alpha}^{(i)} = |\Delta\tilde{\mathbf{t}}^{(i)}| = |\hat{\mathbf{t}}^{(i)} - \hat{\mathbf{t}}^{(i-1)}|, \quad (25)$$

where $\hat{\mathbf{t}}^{(i-1)}$ and $\hat{\mathbf{t}}^{(i)}$ are determined by Kalman filtering.

Using $\hat{\alpha}^{(i)}$, we update $\Delta\tilde{\mathbf{t}}^{(i)}$ (obtained from Eq. 8 using image features and epipolar geometry analysis) to recover absolute translation vector $\Delta\tilde{\mathbf{t}}_*^{(i)}$ between frame $(i-1)$ and i by

$$\Delta\tilde{\mathbf{t}}_*^{(i)} = \hat{\alpha}^{(i)} \frac{\Delta\tilde{\mathbf{t}}^{(i)}}{|\Delta\tilde{\mathbf{t}}^{(i)}|}. \quad (26)$$

Based on our method that integrates SIFT features, epipolar constraints, and Kalman filtering-based scale factor estimation, predicted motion $\Delta\tilde{\mathbf{Q}}_*^{(i)}$ of the bronchoscope camera between frames $(i-1)$ and i can now be formulated as (based on Eq. 1)

$$\Delta\tilde{\mathbf{Q}}_*^{(i)} = \begin{pmatrix} \Delta\tilde{\mathbf{R}}^{(i)} & \Delta\tilde{\mathbf{t}}_*^{(i)} \\ \mathbf{0}^T & 1 \end{pmatrix}, \quad (27)$$

where $\Delta\tilde{\mathbf{R}}^{(i)}$ is calculated by Eq. 9. Finally, estimate $\Delta\tilde{\mathbf{Q}}_*^{(i)}$ is utilized as the initialization of image registration, as described in the next section.

3.3. Intensity-based Image Registration

Intensity-based registration iteratively optimizes the six degrees of freedom describing the position and rotation of a virtual camera located in the 3D CT data. For each camera pose, a 2D VB image can be generated using a ray casting-based volume rendering technique (Mori et al., 2003). In each iteration, this VB image generated from 3D CT is compared to the 2D RB image. The two images are most similar when the virtual and real cameras observe the same physical scene of the patient in the CT and world coordinates.

For calculating the similarity between the two images, several popular schemes are available including mean squared error, normalized cross correlation, mutual information, gradient correlation, and gradient difference (Penney et al., 1998; Viola and Wells, 1997; M.Wells et al., 1996; Hajnal et al., 2001). However, these approaches work effectively and efficiently only if the image similarity has a sharp peak at the camera pose where the two images need to be registered, which is not always the case for bronchoscopic images, particularly when registering VB and RB images that are from different modalities.

Our group developed a selective image similarity measure based on characteristic structures (i.e., fold or bifurcation patterns) observed in bronchoscopic views. This method, which can greatly improve the performance of bronchoscope tracking, consists of four steps: (a) division of the input RB image into subblocks, (b) feature value computation in each subblock, (c) selection of the most appropriate subblocks, and (d) image similarity computation in the selected subblocks. For more details refer to Deguchi et al. (2009).

During intensity-based registration, we use a modified mean squared error (*MoMSE*) similarity measure (Deguchi et al.,

2009) to compute the dissimilarity between RB image $\mathbf{I}_R^{(i)}$ and virtual bronchoscopic (VB) image \mathbf{I}_V . Let $\mathbf{I}_V(\mathbf{Q}^{(i)})$ be a VB image that is generated using rendering parameters $\mathbf{Q}^{(i)} = \mathbf{Q}^{(i-1)}\Delta\mathbf{Q}^{(i)}$. We update inter-frame motion $\Delta\mathbf{Q}^{(i)}$ to generate the most similar VB image $\mathbf{I}_V(\mathbf{Q}^{(i-1)}\Delta\mathbf{Q}^{(i)})$ corresponding to RB image $\mathbf{I}_R^{(i)}$. The complete intensity-based registration process with respect to $\Delta\mathbf{Q}^{(i)}$ can be summarized in the following optimization term

$$\Delta\mathbf{Q}^{(i)} = \arg \min_{\Delta\mathbf{Q}} MoMSE(\mathbf{I}_R^{(i)}, \mathbf{I}_V(\mathbf{Q}^{(i-1)}\Delta\mathbf{Q})). \quad (28)$$

The minimization process is executed using the Powell method (Berghen and Bersini, 2005). In this search process, the initialization of $\Delta\mathbf{Q}$ in Eq. 28 significantly affects the performance of bronchoscope tracking. In Deguchi et al. (2009), $\Delta\mathbf{Q}$ is initialized as an identity matrix. However, in our new hybrid method, we initialize $\Delta\mathbf{Q}$ with our estimate $\Delta\tilde{\mathbf{Q}}_*^{(i)}$ (see Eq. 27) by matching stable image features. Compared to sole image registration, these features are less dependent on airway folds or bifurcations and hence improve the performance of bronchoscope tracking. This is the major advantage of our hybrid method.

4. Experimental Setup

4.1. Dataset Acquisition

To evaluate the performance of all tracking methods, we propose a procedure to acquire bronchoscopic ground truth data (GTD) by combining camera calibration, hand-eye calibration, and electromagnetic tracking. At the beginning of this procedure, we acquire GTD for bronchoscope tracking using an electromagnetic tracking (EMT) system and a rigid airway phantom made of a silicone rubber cast in an acrylic box with 18 bore holes. After attaching a magnetic tracking sensor to the tip of the bronchoscope, we calibrate the intrinsic parameters (including the distortion coefficients) of the bronchoscope camera utilizing Zhang’s method (Zhang, 2000) and estimate the transformation matrix between the camera and sensor coordinate systems by hand-eye calibration (Wengert et al., 2006). Next we register the EMT and CT coordinate frames by taking the 3D measurements of a magnetic sensor inserted into each of the 18 bore holes, identifying the corresponding points in the CT data, and computing the rigid transformation matrix between the two point clouds (Eggert et al., 1997). Afterwards, we insert the magnetically tracked bronchoscope into the phantom and acquire pairs of RB images and EMT outputs. All EMT outputs are transformed into CT coordinates using the calibration and CT-to-EMT registration results and in the following serve as our GTD.

For patient dataset validation, we apply all tracking methods to nine pairs of RB video sequences and 3D CT images based on a standard clinical protocol. The acquisition parameters of the CT images are 512×512 pixels, 72-351 slices, 2.0-5.0-mm slice thickness, and 0.5-2.0-mm reconstruction pitch. Bronchoscopic videos were recorded onto digital videotapes in operation rooms during examinations and transferred to the host

Table 1: Methods evaluated in our experiments. ‘√’ denotes that a particular feature is used by a method. M8 is our proposed hybrid method.

Method		M1	M2	M3	M4	M5	M6	M7	M8
Image Features	Optical Flow	—	—	√	√	√	—	—	—
	SIFT	—	—	—	—	—	√	√	√
Scale Factor	Constant	—	—	√	—	—	√	—	—
	Optimization-Based	—	—	—	√	—	—	√	—
	Kalman-Based	—	—	—	—	√	—	—	√
Kalman Filtering		—	√	—	—	√	—	—	√
Intensity-Based Image Registration		√	√	√	√	√	√	√	√

computer at 30 frames per second. The size of the video frames was 360×370 and 256×263 pixels. We did all implementations on a Microsoft Visual C++ platform and ran the software on a conventional PC (CPU: Intel XEON 3.80 GHz \times 2 processors, 4-GBByte main memory).

4.2. Comparison Methods

Our proposed hybrid method uses SIFT features, epipolar constraints, Kalman-based scale factor determination, and image registration to track the camera motion. To thoroughly evaluate the performance of our new method, we compared it with previously developed state-of-the-art methods. Table 1 summarizes all the methods we evaluated and the details the individual features utilized by each method. While M1 (Deguchi et al., 2009) solely performs image registration to track the camera, M2 (Nagao et al., 2004) additionally applies Kalman filtering to predict the camera motion. M8 corresponds to our proposed method utilizing Kalman filtering to get the unknown scale of the motion matrix obtained by epipolar geometry analysis. To evaluate the performance of the scale factor estimation, we also compared our method to two alternative approaches: (a) In M6, we simply assume the motion is constant and set the scale factor to 0.3, since the bronchoscope movement is of the order of about 0.3 mm/frame (at a video frame rate of 30 fps) (Rai et al., 2008); (b) In M7, we performed nonlinear optimization. Instead of optimizing the full six parameters in Eq. 28, we performed a simple one-parameter intensity-based registration with respect to the scale factor as follows:

$$\check{\alpha}^{(i)} = \arg \min_{\alpha} MoMSE \left(\mathbf{I}_R^{(i)}, \mathbf{I}_V \left(\mathbf{Q}^{(i-1)} \left(\begin{array}{c} \Delta \tilde{\mathbf{R}}^{(i)} \\ \mathbf{0}^T \\ 1 \end{array} \alpha \frac{\Delta \tilde{\mathbf{M}}^{(i)}}{|\Delta \tilde{\mathbf{M}}^{(i)}|} \right) \right) \right). \quad (29)$$

As an alternative to SIFT, optical flow can be utilized to track image features, as proposed in 2002 by Mori et al. (2002) for bronchoscope tracking. The method calculates optical flow for pixels at regular intervals by a simple block matching method that finds the 2D point correspondences between two consecutive RB images. To directly compare and evaluate the performance of SIFT and the optical flow in the context of bronchoscope tracking, our implementation allows the two methods to

be easily interchanged. M3~M5 are the optical flow equivalents of M6~M8.

4.3. Evaluation Criterion

We define the position error as

$$\varepsilon = \|\mathbf{t} - \mathbf{t}_G\|, \quad (30)$$

where ε is the Euclidean distance between reference position \mathbf{t}_G acquired by EMT and estimated position \mathbf{t} obtained by our hybrid tracking method.

For errors in orientation, no generally used measure is defined. As in Schneider and Stevens (2007), we characterize orientation error by the rotation error about the invariant Euler axis:

$$\theta = \arccos((\text{trace}(\mathbf{R}\mathbf{R}_G^T) - 1)/2), \quad (31)$$

where \mathbf{R}_G and \mathbf{R} are the reference and estimated orientation matrices, respectively. We take our EMT-based GTD as reference data.

Additionally, we define a second measure for evaluating whether a method is more robust than another by visually inspecting and counting the number of successfully tracked frames. A frame is tracked successfully if a VB image that is generated from the estimated camera parameters resembles its corresponding RB image. This includes a comparison of the current frame to the previous and successive frames to check whether the tracking result follows the actual motion of the RB camera.

5. Results

5.1. Phantom Assessment

We generated ten phantom GTD sets using the EMT system to evaluate all methods. This phantom study does not involve subject motion, bubbles, and the like. We performed two kinds of evaluations. The first evaluated the accuracy of the motion estimation steps (i.e., excluding the intensity-based image registration step), and the second evaluated the overall tracking performance of the hybrid method including image registration.

Table 2: Examples of position [mm] and orientation error [degrees], always initializing the camera pose of previous frame with GTD instead of actual tracking results and not performing intensity-based registration

Method	GTD Case 1: Frames 3701				GTD Case 2: Frames 3711			
	Position [mm]		Orientation [°]		Position [mm]		Orientation [°]	
	Mean	Std	Mean	Std	Mean	Std	Mean	Std
M2	0.884	0.500			0.682	0.526		
M3	1.538	11.89	0.595	2.318	3.807	27.80	1.167	6.839
M4	1.765	1.150	0.595	2.318	2.860	1.627	1.167	6.839
M5	1.466	11.89	0.595	2.318	3.709	27.81	1.167	6.839
M6	0.945	0.495	0.525	0.543	0.776	0.492	0.458	0.541
M7	1.800	1.187	0.525	0.543	2.816	1.633	0.458	0.541
M8	0.875	0.488	0.525	0.543	0.679	0.512	0.458	0.541

Table 3: Examples of position [mm] and orientation error [degrees], only initializing first frame with GTD and starting continuous tracking

Methods	GTD Case 1: Frames 3701				GTD Case 2: Frames 3711			
	Position [mm]		Orientation [°]		Position [mm]		Orientation [°]	
	Mean	Std	Mean	Std	Mean	Std	Mean	Std
M1	5.279	5.489	17.65	16.29	13.33	13.19	26.68	26.74
M2	13.19	12.05	22.27	15.43	13.30	12.64	27.64	22.72
M3	13.57	13.85	59.53	57.25	11.05	10.93	23.07	18.69
M4	19.47	16.99	19.32	12.22	29.31	24.55	42.12	36.53
M5	3.900	2.227	17.07	11.12	13.72	13.95	30.17	29.36
M6	15.05	16.49	27.42	26.41	4.058	2.227	11.95	6.278
M7	8.447	8.383	21.38	21.67	16.73	15.35	35.00	35.08
M8	2.954	1.029	11.38	4.612	2.742	0.856	9.764	3.774

During the first evaluation, we only evaluated the motion estimation part of all methods, i.e., the capability of Kalman filtering, optical flow, SIFT, Epipolar geometry analysis, and our scale factor estimation to predict camera motion. We (a) use $\mathbf{Q}_G^{(i-1)}$ corresponding to GTD frame $\mathbf{I}_G^{(i-1)}$ to initialize the camera pose corresponding to frame $\mathbf{I}_R^{(i-1)}$, predicted pose $\mathbf{Q}^{(i)}$ of current frame $\mathbf{I}_R^{(i)}$ by one of the presented methods (M2~M8), and (b) continued to use GTD pose $\mathbf{Q}_G^{(i)}$ but not tracking output $\mathbf{Q}^{(i)}$ to initialize next frame $\mathbf{I}_R^{(i)}$, continued tracking frame $\mathbf{I}_R^{(i+1)}$, and so on, repeating (a) and (b). We compute the position and orientation error between $\mathbf{Q}_G^{(i)}$ and tracking output $\mathbf{Q}^{(i)}$ using Eqs. 30 and 31, respectively. In other words, we check how close to GTD the result of our motion estimation methods can get, if they are already initialized by GTD.

Table 2 gives the quantitative results about the performance of all methods for the first two phantom datasets (Cases 1 and 2). The average position error of the SIFT-based methods (M6, M7, and M8) is less than 3 mm, and the average orientation error is not greater than 0.6° . Our accuracy evaluation shows that M8 is the most robust method to estimate both position and orientation. M2 also seems able to accurately predict position information in this nearly ideal setup. Since M2 provides no orientation information, we cannot display the orientation error of M2 in Table 2 and Figure 5.

Figure 5 illustrates that the feature-based motion estimation methods (i.e., optical flow- and SIFT- based methods) give the same results regarding orientation errors within the same group (M3~M5 and M6~M8) because the estimators start with the same orientation matrices. However, we can also see that all SIFT-based methods perform much better than their optical flow counterparts, concluding that SIFT features are more stable and adaptive than optical flow features for image-based bronchoscope tracking.

The second phantom evaluation benchmarks all methods (including the intensity-based image registration step) in terms of position and rotation error compared to GTD and the number of frames correctly tracked. Furthermore, the performances of all the tracking methods are also evaluated based on the number of frames correctly tracked by visual inspection.

Table 3 and Figure 6 show the detailed assessment results for the first two phantom datasets. In contrast to the previous evaluation, here we only initialized the very first frame with GTD. Compared to the other methods, M8 gives the best results. It successfully tracks all frames both in Cases 1 and 2 and has the fewest errors. As for Case 2, M6 can recover tracking after a failure in contrast to the other methods. Figure 7 visualizes the tracking performance for the first two phantom cases and shows examples of RB images and the corresponding VB images generated from the camera parameters calculated by each method.

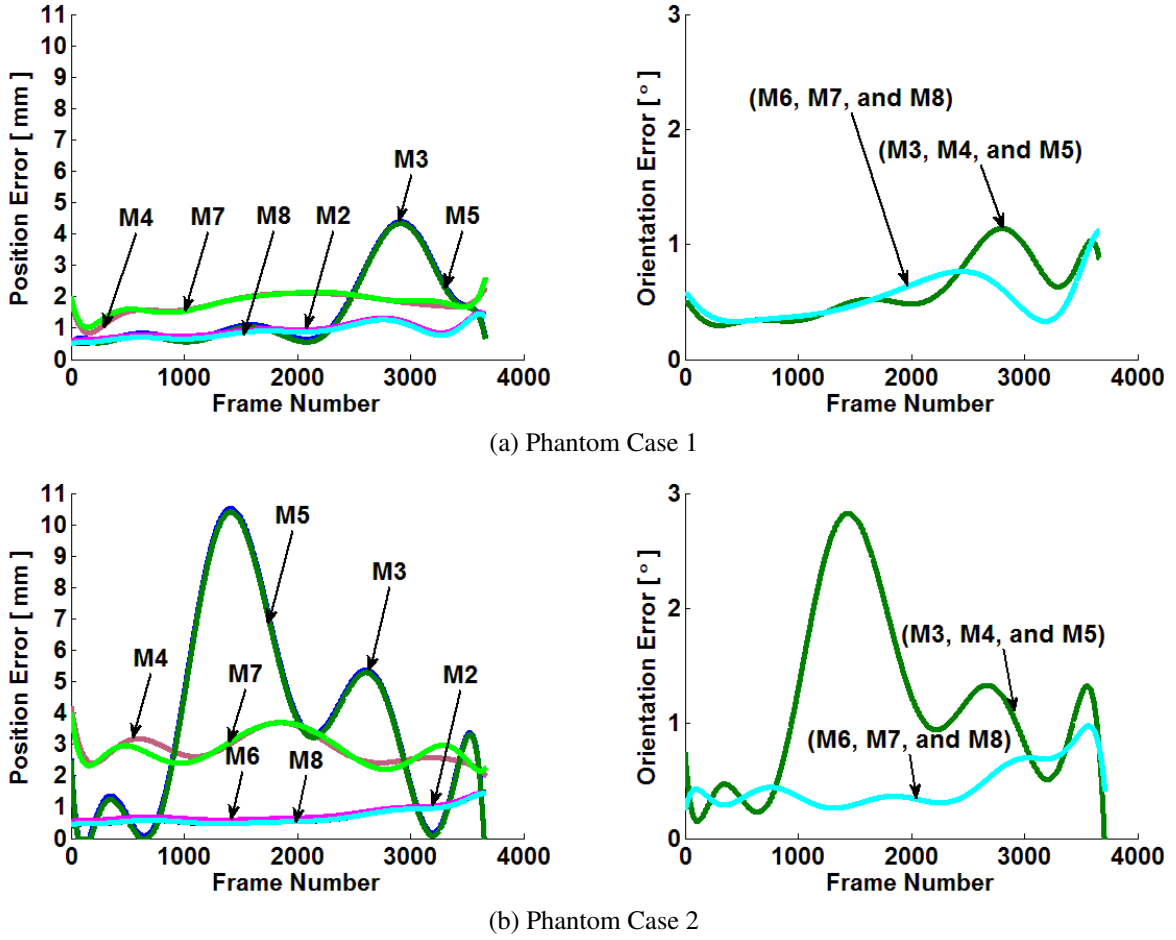


Figure 5: Examples of position and orientation error for different methods when using GTD instead of actual tracking result to initialize pose of previous frame for bronchoscope tracking and without image registration

Table 4 summarizes the quantitative results for all ten phantom datasets. Proposed method M8 significantly improves tracking performance.

Note that the position and orientation errors shown in Tables 2 and 3 are stated relative to GTD. However, GTD itself contains errors stemming from camera and hand-eye calibration, CT-to-EMT registration, and EMT distortions and jitter.

5.2. Patient Assessment

This section reports the performance of our proposed method for human datasets. We applied the methods presented in Sections 3 and 4.2 on patient datasets recorded with acquisition protocols described in Section 4. To evaluate the tracking quality of the eight methods, we again count the number of frames successfully tracked. Note the difficulty of generating GTD for real patients, especially inserting a position sensor into the bronchoscope used in the operating room.

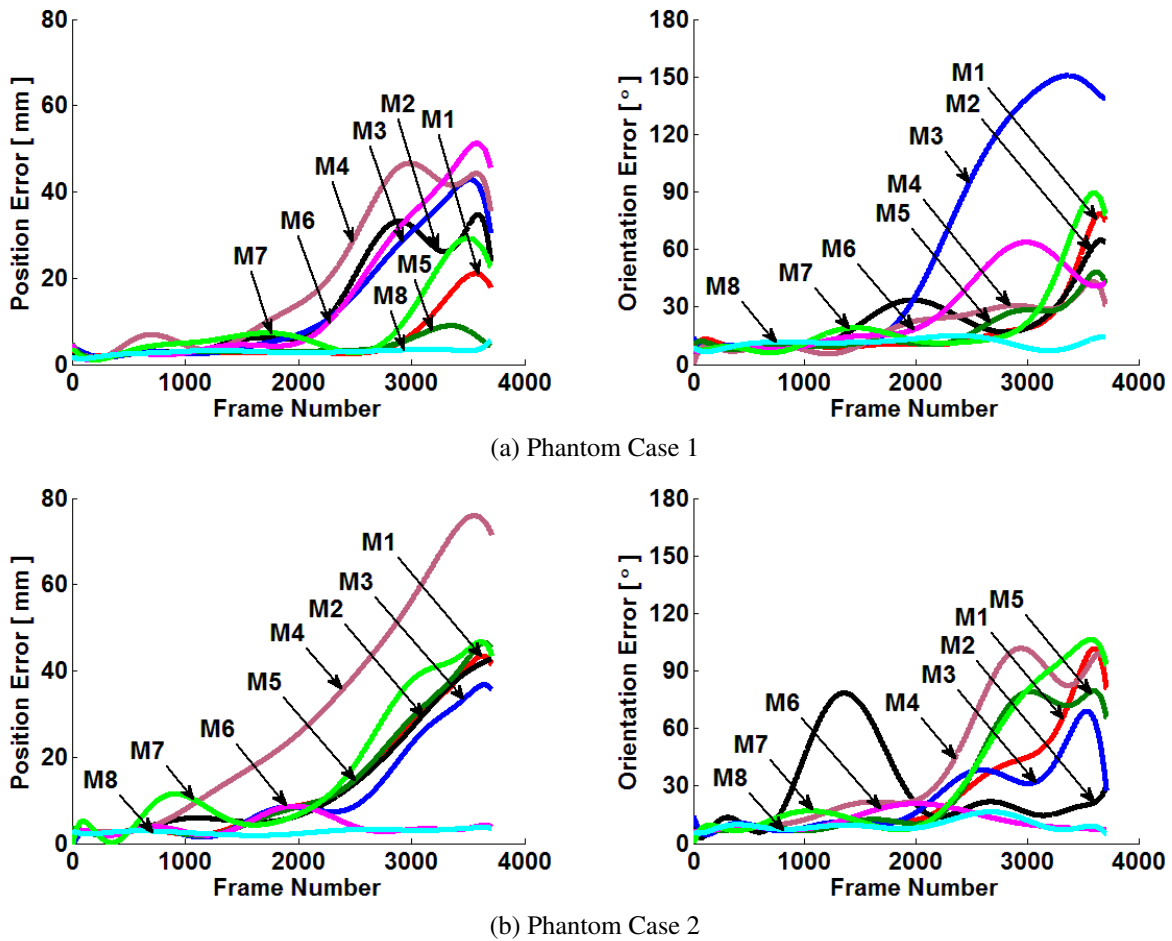
The results are summarized in Table 5 and Figure 9. Table 5 lists the total number of successfully tracked frames for our patient studies and illustrates the effect of intensity-based registration with and without predictive camera tracking on the overall performance. Figure 9 shows examples of the RB images and

the corresponding VB images generated from the camera parameters calculated by each method. For the phantom evaluation results, M8 outperforms all others methods.

6. Discussion

Generally speaking, for both the phantom and patient comparisons it is apparent that in all cases the intensity-based registration accuracy was increased significantly by our proposed hybrid tracking algorithm (method M8).

Note that methods M2~M8 depend on both the results of a rough camera pose estimation step (filtering-based or feature-based motion tracking) and a camera pose refinement step (image registration). If the pose refinement step were omitted, the error would accumulate and tracking of the bronchoscope tip would fail after a few frames. The second image registration step is used to counter such accumulative error. If enough characteristic structure information (e.g., folds and bifurcations) can be collected, the final bronchoscope camera pose can be successfully determined by image registration. If characteristic structure information cannot be observed, the computation of the final bronchoscope camera pose mainly depends on the



(a) Phantom Case 1

(b) Phantom Case 2

Figure 6: Examples of position and orientation error for different methods when using GTD to initialize only first frame for bronchoscope tracking

rough camera pose estimation step. We believe that both steps complement and support each other.

6.1. Phantom Study

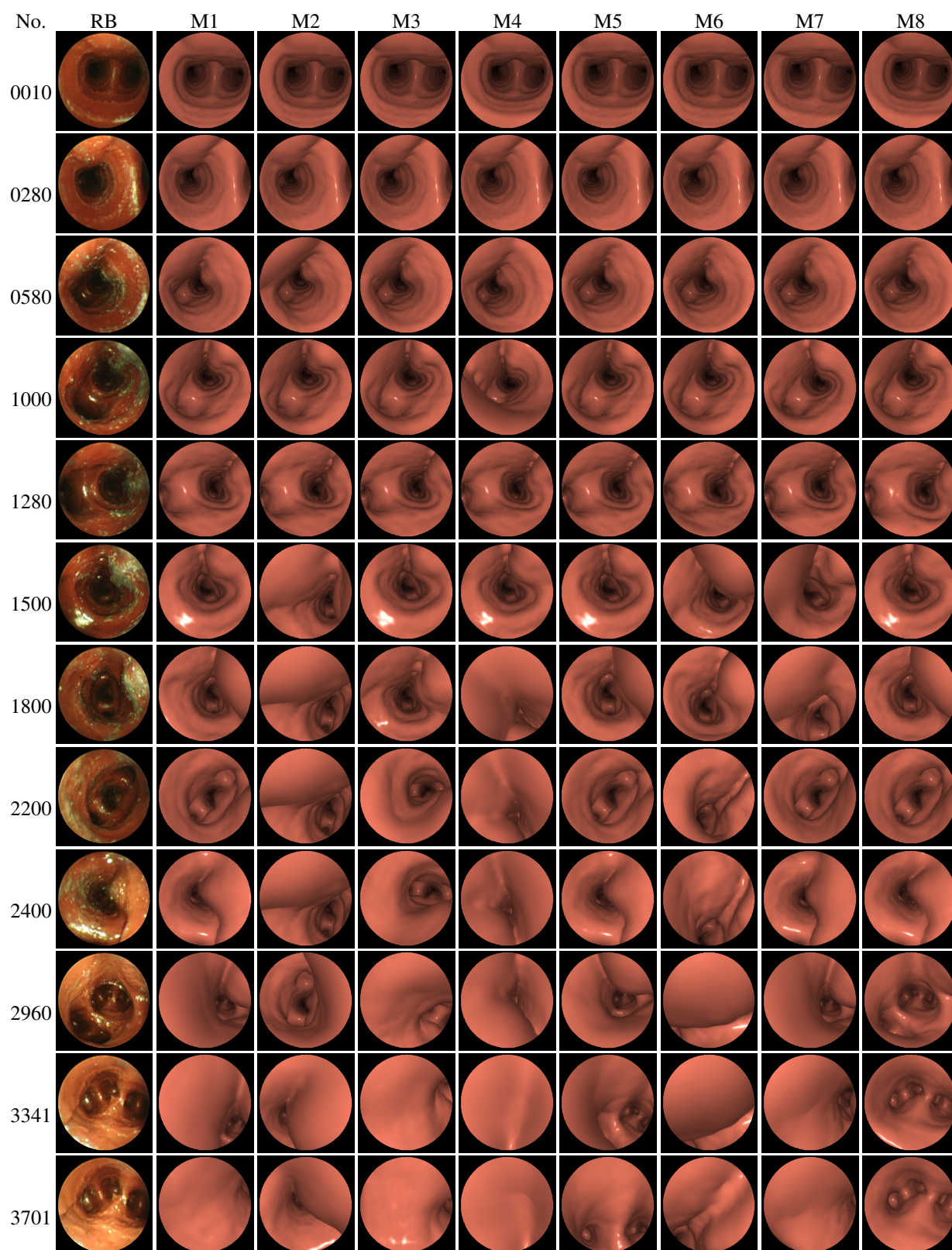
In the phantom study, M8 successfully tracked a total of 16043 bronchoscopic images, increasing the tracking performance by 27.8% more than M1. The tracking performance was particularly improved for Cases 1 and 2 where the bronchoscope took a long path (Figure 7). The intensity-based method sometimes exhibits better tracking performance than the other methods, except M8, because the bronchoscopic images acquired inside the phantom do not have any bubbles, respiration motion, or other deformations. Intensity-based registration alone may collect enough characteristic information to successfully perform registration. This may also be due to epipolar-based camera motion estimation that sometimes gives incorrect initial pose information of the bronchoscope. Since the intensity-based method (M1) instead utilizes the estimation results of the previous frame as initial search parameters, it does not get trapped into local minima.

For future studies we also want to improve GTD accuracy because we used it to initialize the tracking and to measure the

tracking performance during our phantom studies. Basically, obtaining good GTD due to the large errors of the EMT sensor, it is very difficult in particular because the bronchoscope contains conductive metals that distort the magnetic field.

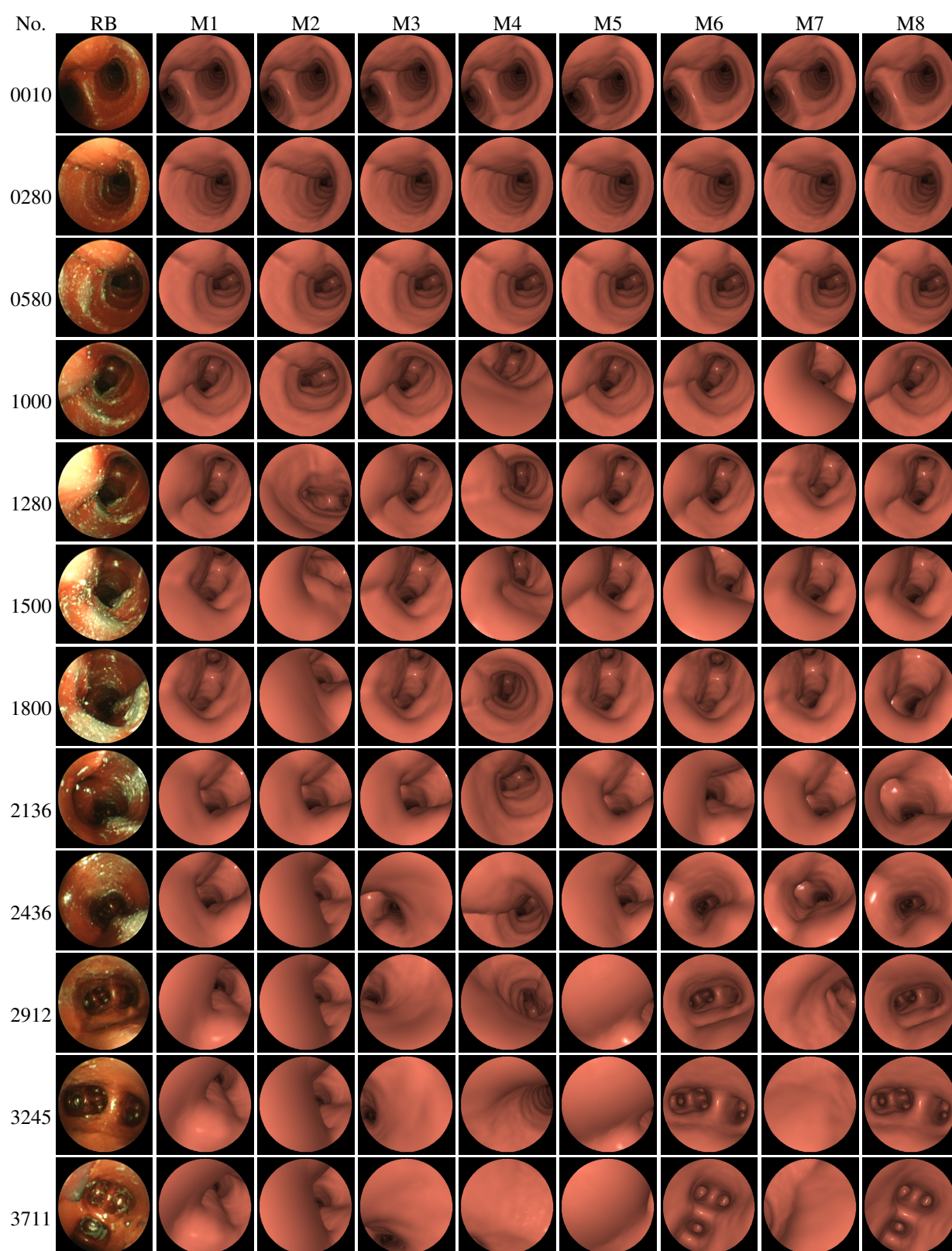
In more detail, GTD accuracy is affected by the following errors: (a) camera and hand-eye calibration error when calculating the transformation between the EMT sensor and the bronchoscope camera, (b) initial registration error when predicting the transformation between the EMT and CT coordinate systems, (c) distortion and jitter of EMT sensor measurements.

The calibration error is mainly caused by: (a) missing the synchronization during the recording of calibration pattern images (shown in Figure 8) and corresponding EMT sensor measurements, and (b) field distortions and jitter affecting the EMT sensor measurements. During calibration, we held the bronchoscope by hand to acquire pairs of pattern images and EMT sensor outputs. This may result in pattern images that are not synchronized to the EMT sensor outputs. In our experiments, we used ten pattern images to perform the hand-eye calibration. The average back-projection error for all pattern images was around 0.5 mm. In the future, we plan to reduce the calibration error by: (a) using a bronchoscope holder to acquire



(a) Phantom Case 1: M8 can successfully track all successive frames

Figure 7: Results of camera motion tracking for phantom assessment. Second column shows selected frames from a sequence of phantom RB images and first column their corresponding frame numbers. Other columns show tracking results for methods M1~M8, all generated by volume rendering of airways from estimated viewpoints.



(b) Phantom Case 2. M8 can successfully track all successive frames.

Figure 7: *Continued.*

Table 4: Quantitative comparison of tracking results for our phantom studies in terms of number and percentage of successfully tracked frames

Phantom Datasets	Num.of Frames	Number and percentage of frames successfully tracked							
		M1	M2	M3	M4	M5	M6	M7	M8
Case 1	3701	2700	1400	2040	1610	2700	1650	2720	3701
		72.9%	37.8%	55.1%	43.5%	72.9%	44.6%	73.5%	100%
Case 2	3711	1600	950	1780	1500	1600	3210	1630	3711
		43.1%	25.6%	47.9%	40.4%	43.1%	84.5%	43.9%	100%
Case 3	1098	760	370	150	100	200	150	400	1012
		69.2%	33.7%	13.7%	9.11%	18.2%	13.7%	36.4%	92.2%
Case 4	1000	608	540	900	550	900	760	900	900
		60.8%	54.0%	90.0%	55.0%	90.0%	76.0%	90.0%	90.0%
Case 5	1400	370	200	920	320	375	312	270	1400
		26.4%	14.3%	65.7%	22.8%	26.8%	22.3%	19.3%	100%
Case 6	2800	1790	1450	1180	600	1620	1740	1000	2240
		63.9%	51.8%	42.1%	21.4%	57.8%	62.1%	35.7%	80.0%
Case 7	522	100	100	300	400	50	150	200	522
		19.1%	19.1%	57.5%	76.6%	9.58%	28.7%	38.3%	100%
Case 8	2200	2100	900	700	390	580	580	560	1500
		95.4%	40.9%	31.8%	17.7%	26.4%	26.4%	25.2%	68.1%
Case 9	1000	568	50	640	100	100	230	200	798
		56.8%	5.00%	64.0%	10.0%	10.0%	23.0%	20.0%	79.8%
Case 10	1500	450	250	400	430	450	600	500	520
		30.0%	16.7%	26.7%	28.7%	30.0%	40.0%	33.3%	34.7%
Total	18932	11046	6210	9010	6000	8575	9382	8360	16304
		58.3%	32.8%	47.6%	31.7%	45.3%	49.6%	44.2%	86.1%

pairs of pattern images and EMT sensor measurements, so precise synchronization is unnecessary and (b) using a mid-range transmitter instead of a flat transmitter for improving EMT accuracy. Currently we use a 3D Guidance medSAFE tracker from the Ascension Technology Corporation with a 4-coil flat transmitter as a magnetic field generator and EMT sensors with diameters of only 1.3 mm. As previously shown EMT sensor outputs can be more accurate if a mid-range transmitter is used instead of a flat transmitter because the former has higher excitation than the latter (Feuerstein et al., 2009).

When performing point-based rigid registration to estimate the transformation between the EMT and CT coordinate systems (Eggert et al., 1997; Solomon et al., 2000; Becker et al., 2005), we achieved registration error of about 1.2 mm. This error stems from two major sources. One is that we had to manually choose the landmarks in the CT coordinate system, since they are drill holes in the phantom. For our next phantom, we will use spheres attached to the phantom that can be easily and automatically segmented in CT images. The other is that in the EMT coordinate system they are again manually collected by inserting an EMT sensor into the drill holes. The sensor measurements themselves only have limited static accuracy and resolution. This limitation of EMT remains challenging. It is not easy to reduce the error of an EMT system that is usually com-

prised of two kinds of error: (a) static, and (b) dynamic. While static error can be partially corrected (Kindratenko, 2000), dynamic error caused, e.g., by ferromagnetic material (which also contains the bronchoscope) is difficult to correct, unless combined with optical tracking (Feuerstein et al., 2009).

Generally, we plan to use more accurate EMTs and more elaborate calibration setup when we evaluate any new methods for bronchoscope tracking. This will be future work.

6.2. Patient Study

In the patient comparison, the robustness of our tracking methods decreased because of bronchoscopic video frames problematic due to coughing, the appearance of bubbles, deformation caused by complicated breathing patterns, the bronchoscope viewing the bronchial surface, or quick insertion or extraction of the bronchoscope. However, our proposed approach M8 still shows better tracking performance than all other methods. For example, in Case K and Path 2 of Case A, bubbles, large bronchi deformation, and motion blur were observed in the RB images. Except M8, all methods failed to appropriately track bronchoscopic motion. We see two main reasons behind this improvement of M8. First, SIFT feature-based estimation gives an accurate rotation matrix and translation vectors. Second, Kalman-based scale calculation for camera motion pre-

Table 5: Quantitative comparison of tracking results for our patient studies in terms of number and percentage of successfully tracked frames

Cases (Path)	Num. of Frames	Number and percentage of frames successfully tracked							
		M1	M2	M3	M4	M5	M6	M7	M8
Case A(1)	800	618	772	690	730	690	800	800	774
		77.2%	96.5%	86.2%	91.2%	86.2%	100%	100%	96.7%
Case A(2)	800	340	45	340	340	300	340	340	780
		42.5%	5.63%	42.5%	42.5%	37.5%	42.5%	42.5%	97.5%
Case B	500	143	82	320	380	120	100	90	450
		28.6%	16.4%	64.0%	76.0%	24.0%	20.0%	18.0%	90.0%
Case C(1)	1000	1000	800	935	793	1000	1000	1000	1000
		100%	80.0%	93.5%	79.3%	100%	100%	100%	100%
Case C(2)	500	230	20	160	160	500	170	180	180
		46.0%	4.00%	32.0%	32.0%	100%	34.0%	36.0%	36.0%
Case D	429	429	100	386	210	420	200	200	300
		100%	23.3%	90.0%	48.9%	97.9%	46.6%	46.6%	69.9%
Case E(1)	993	993	350	900	800	993	993	400	993
		100%	35.2%	90.6%	80.6%	100%	100%	40.3%	100%
Case E(2)	500	320	350	420	320	420	420	420	450
		64.0%	70.0%	84.0%	64.0%	84.0%	84.0%	84.0%	90.0%
Case E(3)	1000	260	540	460	530	590	370	520	750
		26.0%	54.0%	46.0%	53.0%	59.0%	37.0%	52.0%	75.0%
Case E(4)	279	279	200	279	279	279	279	279	279
		100%	71.7%	100%	100%	100%	100%	100%	100%
Case F	1000	130	130	124	125	140	130	180	260
		13.0%	13.0%	12.4%	12.5%	14.0%	13.0%	18.0%	26.0%
Case G	450	10	10	240	10	240	430	180	420
		2.22%	2.22%	53.3%	2.22%	53.3%	95.5%	40.0%	93.3%
Case H	1200	140	20	60	80	100	1200	100	140
		11.7%	1.67%	5.00%	6.67%	8.33%	100%	8.33%	11.7%
Case K	1200	560	560	180	150	220	690	150	1120
		46.7%	46.7%	15.0%	12.5%	18.3%	57.5%	12.5%	93.3%
Total	10651	5452	3979	5494	4907	6012	7122	4839	7916
		51.2%	37.4%	51.6%	46.1%	56.4%	66.9%	45.4%	75.0%

diction contributes to this improvement. Kalman filtering can in most cases properly estimate the correct depth information between consecutive frames. The Kalman-based scale determination method is generally better than the other two methods because the optimization-based method depends on the image registration result, which in turn depends on characteristic structures such as folds or bifurcations. However, if we obtain similar 2D VB images for different frames due to missing characteristic structures, we may end up with a scale factor close or equal to 0, even though the two frames were taken from different RB camera positions. Moreover, for the method that assumes the scale factor to be constant, this hard constraint obviously cannot be valid for all frames, since the camera does not move at a constant motion.

Additionally, note that Kalman filtering is used in its sim-

plest form in all Kalman-based methods. We used exactly the same implementation as in Nagao et al. (2004) for fair comparisons. Its main restriction is the linearity of the model. We may improve the performance using extended Kalman filtering (EKF) (Welch and Bishop, 2001), which is a non-linear model of Kalman filtering. During the implementation of Kalman filtering, we model the observation noise to be zero and filter our measurements directly, since adding observation noise did not improve the results. The tracking results are not affected either when changing the covariance matrix settings from our testing experience. For example, when we set the covariance of the white noise to be $\mathbf{W}_{i-1} = 0.005 \cdot \mathbf{I}_{9 \times 9}$ or $\mathbf{W}_{i-1} = 0.5 \cdot \mathbf{I}_{9 \times 9}$, the tracking performance did not change. The only change was an increase in the computation time. This means filtering took “shorter or longer to believe” the measurements (Welch

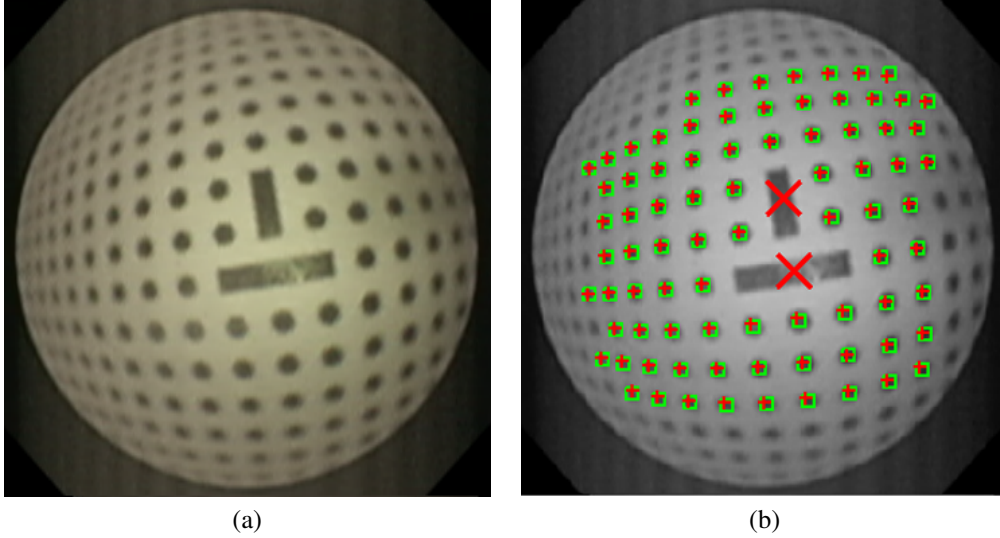


Figure 8: Example of pattern image used for camera calibration and its back-projection error: (a) recorded pattern image and (b) back-projection image after calibration. Green squares are real (segmented) positions, and red crosses are estimated positions.

and Bishop, 2001). Otherwise, we must clarify that it may not be reasonable to assume invariant acceleration in movements when using Kalman filtering. This assumption can only be satisfied when force \mathbf{F} , moving the bronchoscope of mass \mathbf{m} , is constant, because $\mathbf{F} = \mathbf{m}\mathbf{a}$, with acceleration \mathbf{a} . This may only be true in the case when bronchoscopists keep constant force to continuously move the bronchoscope.

There are several cases in which the performance of some methods was almost the same as that of the intensity-based method (M1), such as Cases 10 and F. This is because the estimation failed in the intensity-based registration process, even if predictive tracking gave good initialization. Therefore, additional future work will address the performance of intensity-based image registration.

6.3. Computational Efficiency

Our proposed method M8 requires approximately 1.8 seconds to process a frame. This is because feature-based motion recovery methods are time-consuming in terms of the detection of points and finding correspondences, and so is the registration stage of bronchoscope tracking. However, our future work will utilize GPU to accelerate our implementations and approximate real time, as already shown independently for SIFT (Wu, 2007) and intensity-based image registration (Sugiura et al., 2009).

Figure 10 shows examples of computation times using methods M1, M2, and M8. The detection of SIFT features is time-consuming, particularly if implemented on CPU.

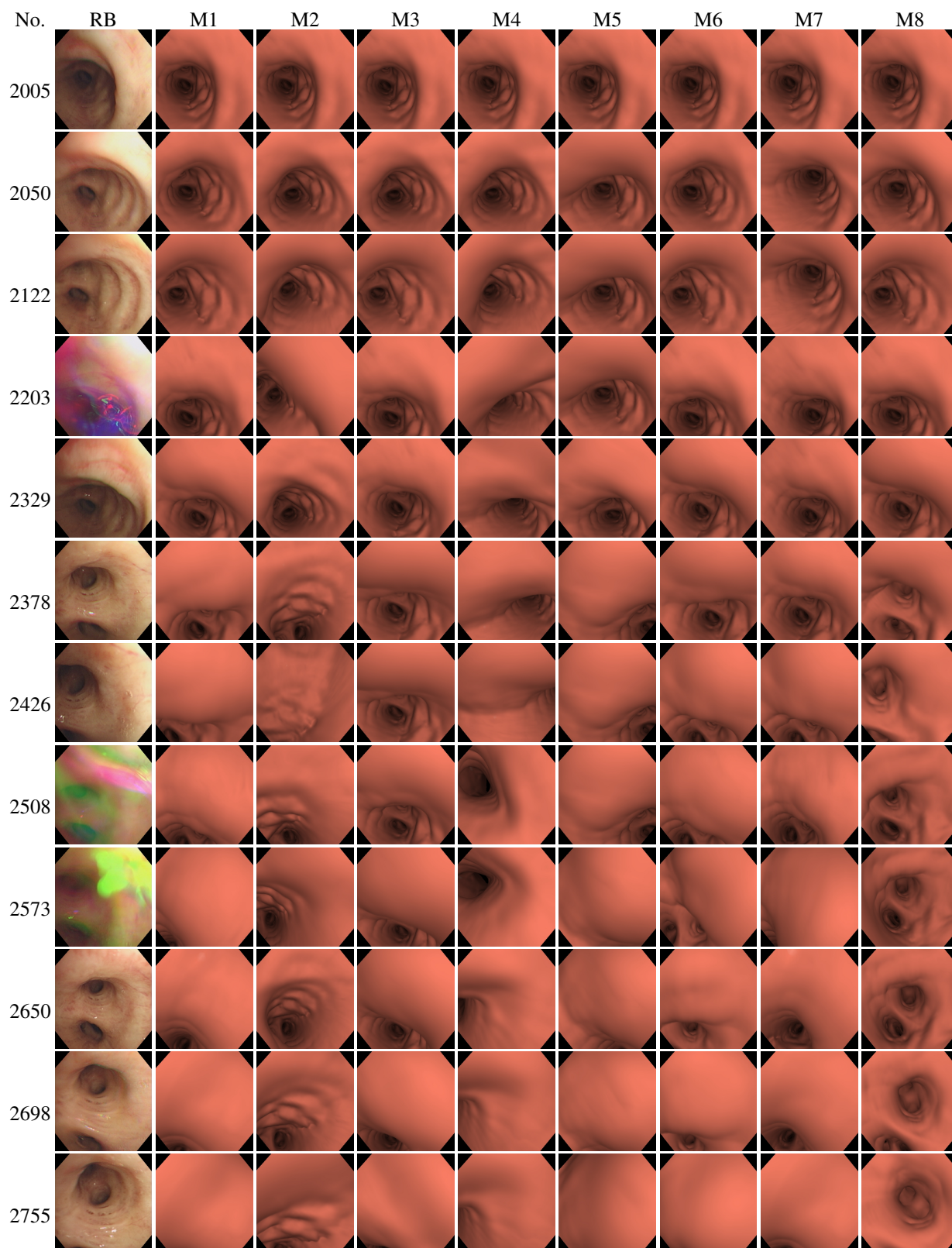
6.4. Challenges

All the methods presented here fail to continuously track the bronchoscope in some cases (e.g., Cases 10 and F). Apart from problematic bronchoscope video frames that cause tracking failure, we made several assumptions that in practice do not always hold.

First, we suppose that our feature-based motion estimation can always detect good and stable features and find correct correspondences. However, it is sometimes difficult to find reliable point correspondences between images, especially for large camera movements and (dis)appearing bubbles among successive frames. For example, in Cases C(2) and D, it is not easy to find corresponding SIFT features because many bubbles appear and disappear among successive images, and therefore camera pose estimation cannot obtain good initialization for image registration.

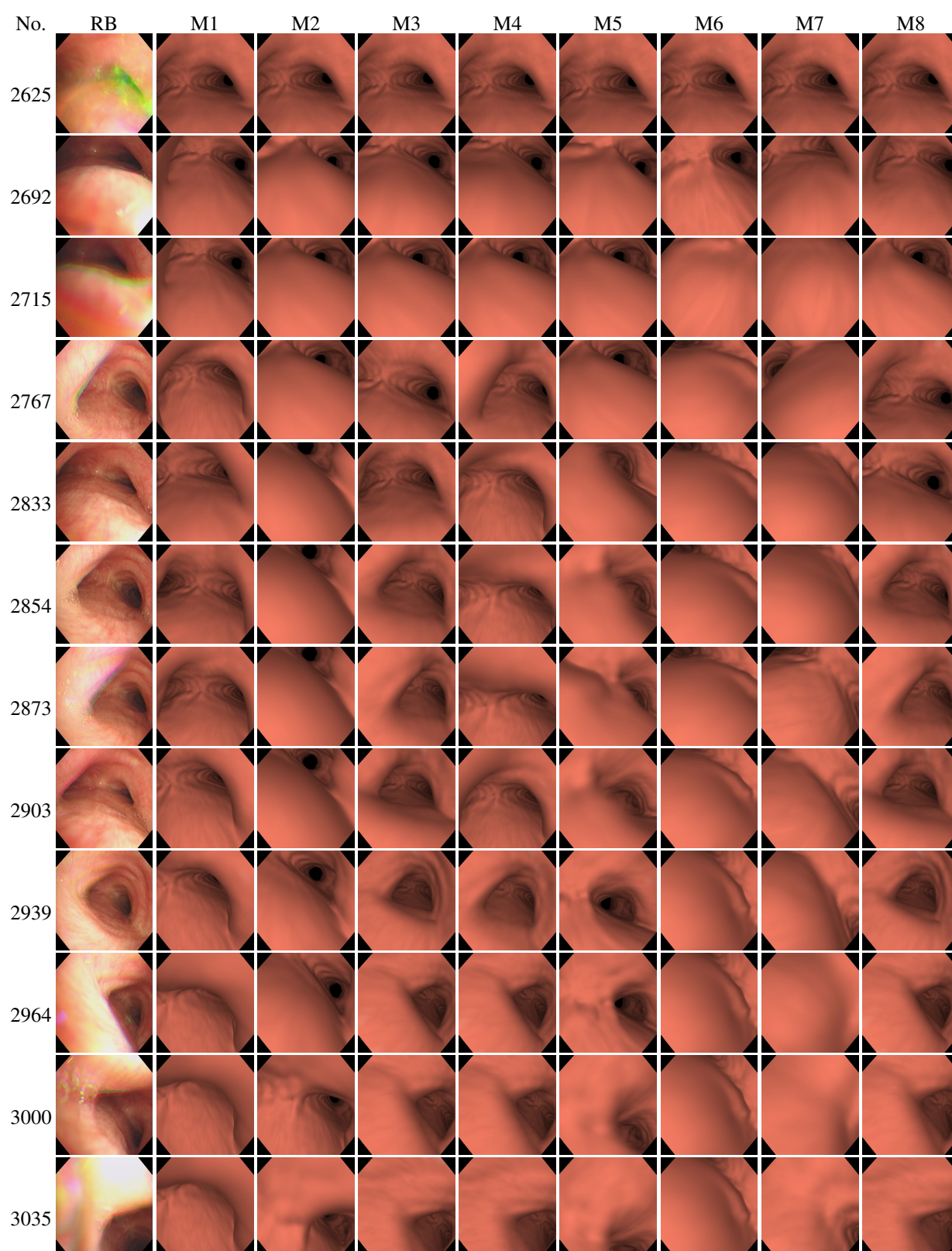
Second, we assume that a real bronchoscope camera is in a constant state of acceleration. This is a hard constraint since, in practice, acceleration can change quickly. Our motion recovery method cannot properly predict when the camera moves or changes direction quickly, particularly for the Kalman-based method. This may explain why setting the scale to a constant sometimes shows better tracking results than using Kalman-based scale determination. For example, in Case H, the camera changes its direction of movement too quickly and stops entirely two times. This situation definitely violates the assumption of Kalman filtering. This means that our method fails whenever there is large mis-synchronization. In future work, we plan to measure the insertion depth of the bronchoscope to solve this problem.

Third, airways deformation caused by patient movement, breathing, and coughing is one particular challenge in navigated bronchoscopy. Currently we do not explicitly model respiratory motion in our tracking method. In general, we assume that image registration-based tracking is not as sensitive to airways deformation as EMT, because standard EMT approaches just assume a global rigid transformation between the CT and EMT coordinate systems, and image registration locally registers VB and RB images. However, in Case B, large airway deformation can be observed. Although our method can track bronchoscope motion for most frames, it still cannot precisely



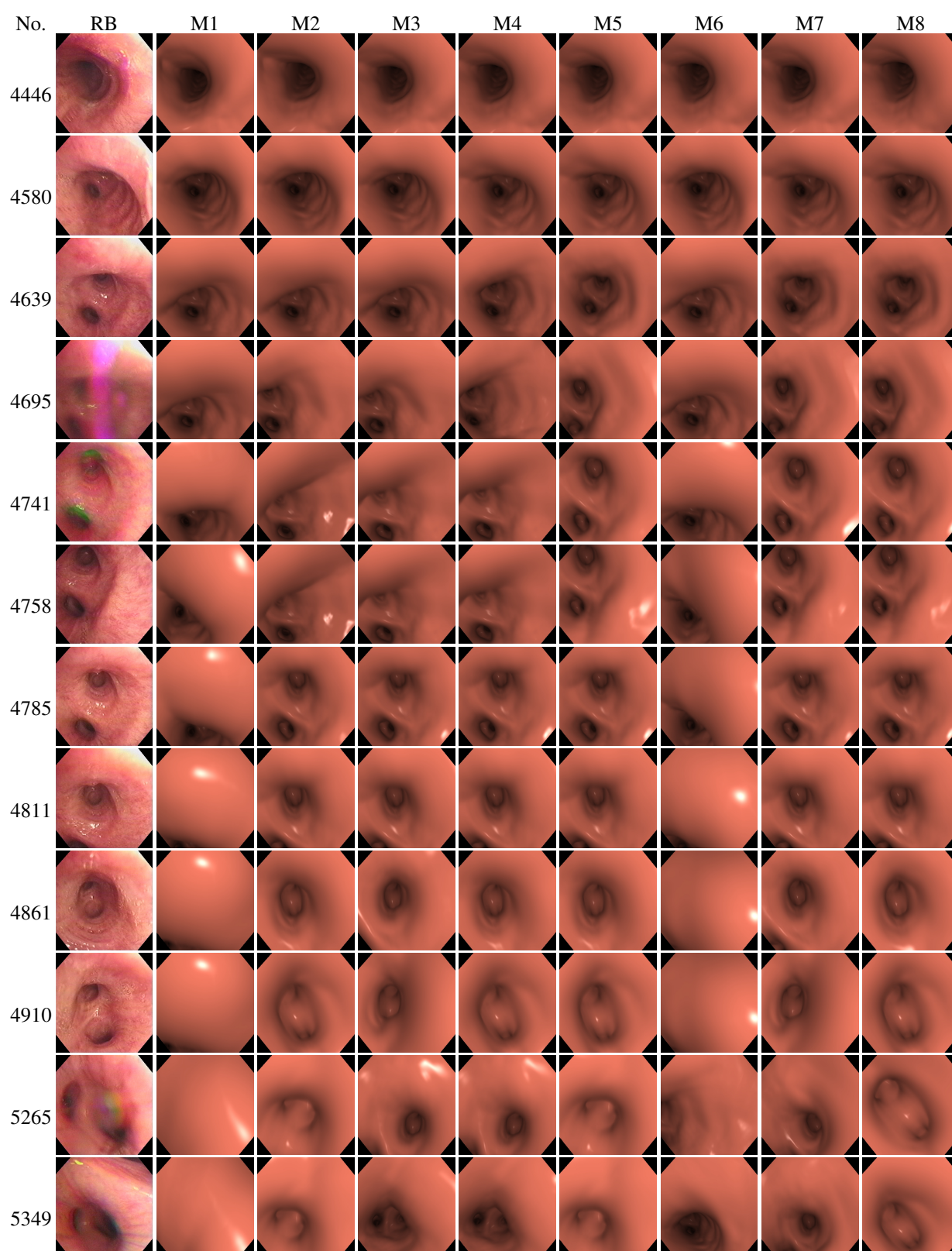
(a) Path 2 of Case A: M8 successfully tracks most frames

Figure 9: Results of camera motion tracking for patient assessment. Second column shows selected frames from a sequence of patient RB images and first column their corresponding frame numbers. Other columns show tracking results for methods M1~M8, all generated by volume rendering of airways from estimated viewpoints.



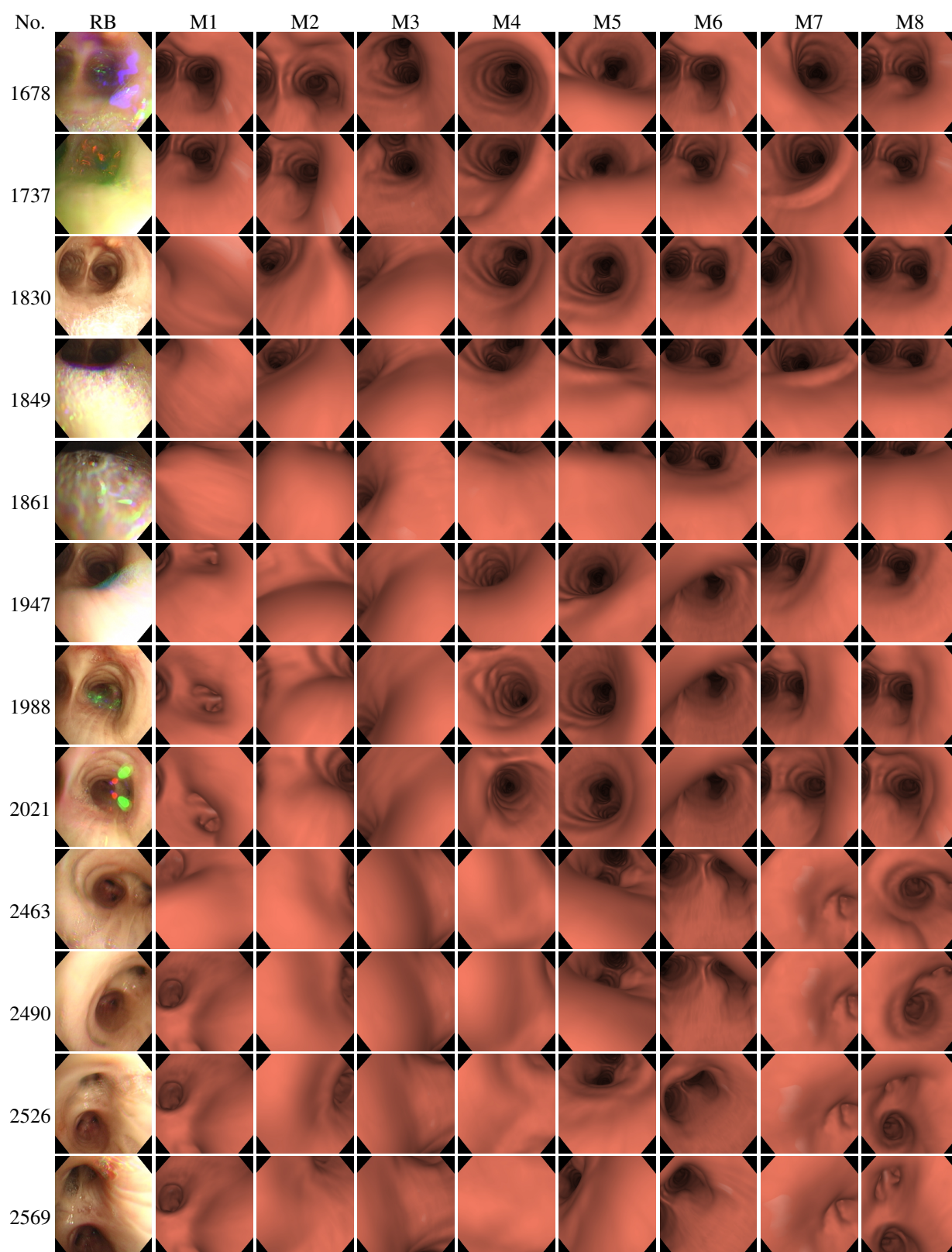
(b) Case B: M8 successfully tracks most frames

Figure 9: *Continued.*



(c) Path 3 of Case E: M8 successfully tracks most frames

Figure 9: *Continued.*



(d) Case K: M8 successfully tracks most frames

Figure 9: *Continued.*

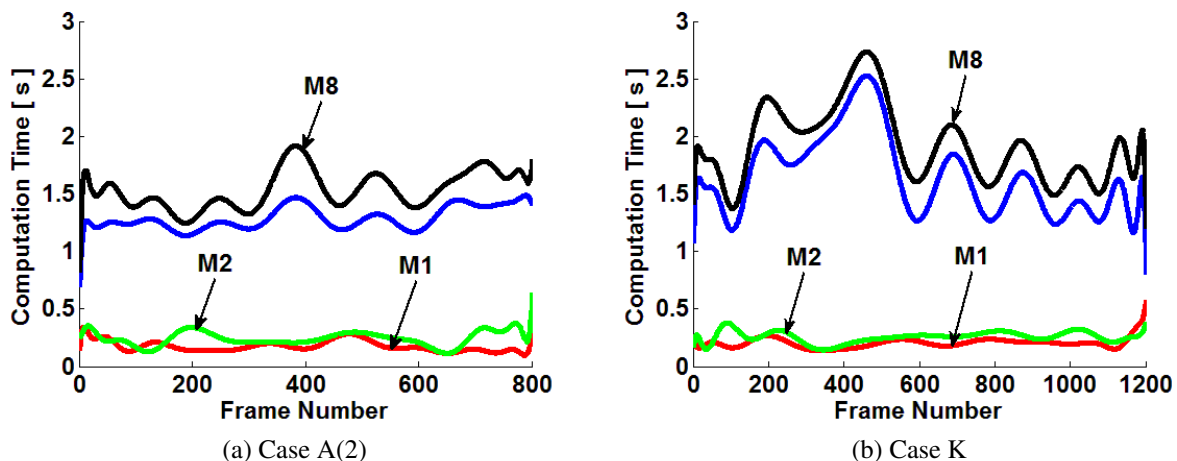


Figure 10: Examples of computation times using methods M1, M2, and M8 for Cases A(2) and K. It clearly shows that for M8 average processing time with at least 1.5 seconds per frame is much higher than that for M1 and M2, because it includes SIFT feature detection. Blue line only shows time required for detecting SIFT features.

register RB and VB images because VB images are generated from CT data acquired during patient breath holds in either inspiration or expiration. We have to compensate for respiratory motion to improve the performance of bronchoscope tracking in the future.

Fourth, we assume that the real world 3D space seen by the bronchoscope coincides with the virtual world 3D space represented by CT. This implies that the entire bronchial structure (characteristic parts) observed in the RB images is surmised to also be visible in VB images at any time during registration. However, the bronchial structure is not well reproduced in VB images due to the low resolution of the CT images. Furthermore, when the CT data of patients are acquired, they are usually asked to fully inhale or exhale and hold their breath. However, during bronchoscopy they are breathing regularly, deforming the airways.

Fifth, we suppose that we can accurately initialize the first VB camera pose in the CT coordinates corresponding to their RB pose at the beginning of the bronchoscope tracking. In our phantom studies, we initialize the pose of the first frame using ground truth data obtained by EMT. In our patient studies, we manually adjusted the six degrees of freedom of the virtual camera inside the 3D CT volume to generate an initial 2D VB image that resembles the corresponding RB image using a graphical user interface. In the operating room, we may integrate this graphical user interface into the bronchoscopic navigation system. However, since manual initialization is time-consuming for a bronchoscopist, we also plan to accomplish initialization by adding an automatic camera position estimation algorithm (Shinohara et al., 2006; Merritt et al., 2006).

7. Conclusion

In this paper, we presented a new hybrid motion tracking method for bronchoscopic navigation and compared it to other

state-of-the-art methods in a comprehensive evaluation conducted on phantom and patient datasets. Our experimental results show that our hybrid method based on SIFT features, epipolar geometry analysis, Kalman filtering, and image registration significantly improves the tracking performance compared to other evaluated methods. Moreover, we do not depend on an additional electromagnetic tracking sensor. Our hybrid method is promising for bronchoscope tracking during bronchoscopic navigation.

Acknowledgments

The authors would like to thank Dr. Y. Hasegawa from the Graduate School of Medicine, Nagoya University, Japan, for his support. This work was partly supported by the “China Scholarship Council Postgraduate Scholarship Program,” by the JSPS postdoctoral fellowship program for foreign researchers, by the program of the formation of innovation center for the fusion of advanced technologies called the “Establishment of early preventing medical treatment based on medical-engineering for analysis and diagnosis” funded by MEXT, by a Grant-in-Aid for Science Research funded by MEXT and JSPS, and by a Grant-in-Aid for Cancer Research funded by the Ministry of Health, Labour and Welfare, Japan. The authors are grateful to the anonymous reviewers for their many helpful suggestions, which have improved this paper significantly.

References

- Becker, H. D., Herth, F., Ernst, A., Schwarz, Y., 2005. Bronchoscopic biopsy of peripheral lung lesions under electromagnetic guidance: a pilot study. *Journal of Bronchology & Interventional Pulmonology* 12 (1), 9–13.
- Berghen, F. V., Bersini, H., 2005. CONDOR, a new parallel, constrained extension of powell’s UOBYQA algorithm: experimental results and comparison with the DFO algorithm. *Journal of Computational and Applied Mathematics* 181 (1), 157–175.

- Bricault, I., Ferretti, G., Cinquin, P., 1998. Registration of real and CT-derived virtual bronchoscopic images to assist transbronchial biopsy. *IEEE Transactions on Medical Imaging* 17 (5), 703–714.
- Chung, A. J., Deligianni, F., Shah, P., Wells, A., Yang, G. Z., 2006. Patient-specific bronchoscopy visualization through BRDF estimation and disocclusion correction. *IEEE Transactions on Medical Imaging* 25 (4), 503–513.
- Davison, A. J., Reid, I. D., Molton, N. D., Stasse, O., 2007. MonoSLAM: real-time single camera SLAM. *IEEE Transaction on Pattern Analysis and Machine Intelligence* 29 (6), 1052–1067.
- Deguchi, D., Mori, K., Feuerstein, M., Kitasaka, T., Maurer Jr., C. R., Suenaga, Y., Takabatake, H., Mori, M., Natori, H., 2009. Selective image similarity measure for bronchoscope tracking based on image registration. *Medical Image Analysis* 13 (4), 621–633.
- Deligianni, F., Chung, A., Yang, G. Z., 2005. Predictive camera tracking for bronchoscope simulation with CONDensation. In: *Proceedings of MICCAI*. Vol. 3749. pp. 910–916.
- Deligianni, F., Chung, A. J., Yang, G. Z., 2006. Nonrigid 2-D/3-D registration for patient specific bronchoscopy simulation with statistical shape modeling: Phantom validation. *IEEE Transactions on Medical Imaging* 25 (11), 1462–1471.
- Durrant-Whyte, H., Balley, T., 2006. Simultaneous localization and mapping: part I. *IEEE Robotics & Automation Magazine* 13 (2), 99–108.
- Eggert, D., Lorusso, A., Fisher, R., 1997. Estimating 3-D rigid body transformations: a comparison of four major algorithms. *Machine Vision and Applications* 9 (5-6), 272–290.
- Feuerstein, M., Reichl, T., Vogel, J., Traub, J., Navab, N., June 2009. Magneto-optical tracking of flexible laparoscopic ultrasound: Model-based online detection and correction of magnetic tracking errors. *IEEE Transactions on Medical Imaging* 28 (6), 951–967.
- Gildea, T. R., Mazzone, P. J., Karnak, D., Meziane, M., Mehta, A. C., 2006. Electromagnetic navigation diagnostic bronchoscopy: a prospective study. *American Journal of Respiratory and Critical Care Medicine* 174 (9), 982–989.
- Grasa, O. G., Civera, J., Güemes, A., Muñoz, V., Montiel, J., 2009. EKF monocular SLAM 3d modeling, measuring and augmented reality from endoscope image sequences. In: *5th Workshop on Augmented Environments for Medical Imaging including Augmented Reality in Computer-Aided Surgery*.
- Hajnal, J. V., Hill, D. L. G., Hawkes, D. J., 2001. *Medical image registration*. CRC Press.
- Hartley, R., Zisserman, A., 2004. *Multiple view geometry in computer vision*. Cambridge University Press.
- Heikkilä, J., Silvén, O., 1997. A four-step camera calibration procedure with implicit image correction. In: *Proceedings of IEEE CVPR 1997*. pp. 1106–1112.
- Helferty, J., Sherbondy, A., Kiraly, A., Higgins, W., 2007. Computer-based system for the virtual-endoscopic guidance of bronchoscopy. *Computer Vision and Image Understanding* 108, 171–187.
- Helferty, J. P., Higgins, W. E., 2001. Technique for registering 3D virtual CT images to endoscopic video. In: *Proceedings of ICIP (International Conference on Image Processing)*. pp. 893–896.
- Kalman, R. E., 1960. A new approach to linear filtering and prediction problems. *Transaction of the ASME Journal of Basic Engineering* 82 (Series D), 35–45.
- Kindratenko, V. V., 2000. A survey of electromagnetic position tracker calibration techniques. *Virtual Reality* 5 (3).
- Lowe, D. G., 2004. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision* 60 (2), 91–110.
- Merritt, S. A., Rai, L., Higgins, W. E., 2006. Real-time CT-video registration for continuous endoscopic guidance. In: *Proceedings of SPIE*. Vol. 6143. pp. 614313–1–15.
- Mikolajczyk, K., Schmid, C., 2005. A performance evaluation of local descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 27 (10), 1615–1630.
- Mori, K., Deguchi, D., Akiyama, K., Kitasaka, T., Maurer Jr., C. R., Suenaga, Y., Takabatake, H., Mori, M., Natori, H., 2005. Hybrid bronchoscope tracking using a magnetic tracking sensor and image registration. In: *Proceedings of MICCAI 2005*. Vol. LNCS 3750. pp. 543–550.
- Mori, K., Deguchi, D., Sugiyama, J., Suenaga, Y., Toriwaki, J., Maurer Jr., C. R., Takabatake, H., Natori, H., 2002. Tracking of a bronchoscope using epipolar geometry analysis and intensity based image registration of real and virtual endoscopic images. *Medical Image Analysis* 6, 321–336.
- Mori, K., Suenaga, Y., Toriwaki, J., 2003. Fast software-based volume rendering using multimedia instructions on PC platforms and its application to virtual endoscopy. In: *Proceedings of SPIE*. Vol. 5031. pp. 111–122.
- M.Wells, W., Viola, P., Atsumi, H., Nakajima, S., Kikinis, R., 1996. Multimodal volume registration by maximization of mutual information. *Medical Image Analysis* 1, 35–61.
- Nagao, J., Mori, K., Enjouji, T., Deguchi, D., Kitasaka, T., Suenaga, Y., Hasegawa, J., Toriwaki, J., Takabatake, H., Natori, H., 2004. Fast and accurate bronchoscope tracking using image registration and motion prediction. In: *Proceedings of MICCAI*. Vol. 3217. pp. 551–558.
- Penney, G. P., Weese, J., Little, J. A., Desmedt, P., Hill, D. L. G., Hawkes, D. J., 1998. A comparison of similarity measures for use in 2-D-3-D medical image registration. *IEEE Transactions on Medical Imaging* 17 (4), 175–185.
- Peters, T., Cleary, K., 2008. *Image-guided interventions: technology and applications*. Springer, Berlin.
- Rai, L., Helferty, J. P., Higgins, W. E., 2008. Combined video tracking and image-video registration for continuous bronchoscopic guidance. *International Journal of Computer Assisted Radiology and Surgery* 3 (3-4), 315–329.
- Schneider, M., Stevens, C., 2007. Development and testing of a new magnetic-tracking device for image guidance. In: *Proceedings of SPIE*. Vol. 6509. p. 65090L.
- Schwarz, Y., Greif, J., Becker, H. D., Ernst, A., Mehta, A., 2006. Real-time electromagnetic navigation bronchoscopy to peripheral lung lesions using overlaid CT images: the first human study. *Chest* 129 (4), 988–994.
- Shinohara, R., Mori, K., Deguchi, D., Kitasaka, T., Suenaga, Y., Takabatake, H., Mori, M., Natori, H., 2006. Branch identification method for CT-guided bronchoscopy based on eigenspace image matching between real and virtual bronchoscopic images. In: *Proceedings of SPIE*. Vol. 6143. pp. 614314–1–12.
- Solomon, S. B., P. White, J., Wiener, C. M., Orens, J. B., Wang, K. P., 2000. Three-dimensional CT-guided bronchoscopy with a real-time electromagnetic position sensor: a comparison of two image registration methods. *Chest* 118 (6), 1783–1787.
- Sugiura, T., Deguchi, D., Feuerstein, M., Kitasaka, T., Suenaga, Y., Mori, K., 2009. A method for accelerating bronchoscope tracking based on image registration by using GPU. In: *Proceedings of SPIE*. Vol. 7261. pp. 726108–1–12.
- Viola, P., Wells, W. M., 1997. Alignment by maximization of mutual information. *International Journal of Computer Vision* 24 (2), 37–154.
- Welch, G., 1996. SCAAT: Incremental tracking with incomplete information. Ph.D. thesis, University of North Carolina at Chapel Hill, USA.
- Welch, G., Bishop, G., 2001. An introduction to the kalman filter. In: *SIG-GRAPH 2001*, Course 8.
- Wengert, C., Reeff, M., Cattin, P. C., Szekely, G., 2006. Fully automatic endoscope calibration for intraoperative use. In: *Bildverarbeitung für die Medizin*, Springer-Verlag. Vol. 19-21. pp. 419–423.
- Wu, C. C., 2007. SiftGPU: A GPU implementation of scale invariant feature transform (SIFT). <http://cs.unc.edu/~ccwu/siftgpu>.
- Zhang, Z., 2000. A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22 (11), 1330–1334.