Published in final edited form as:

IEEE Trans Med Imaging. 2013 August ; 32(8): 1376–1396. doi:10.1109/TMI.2013.2252361.

Interactive CT-Video Registration for the Continuous Guidance of Bronchoscopy

Scott A. Merritt,

X-Nav Technologies LLC, Lansdale, PA 19446 USA

Rahul Khare,

Sheikh Zayed Institute at the Childrens National Medical Center, Washington, DC 20010 USA

Rebecca Bascom, and

Department of Medicine, Pennsylvania State Hershey Medical Center, Hershey, PA 17033 USA

William E. Higgins [Fellow, IEEE]

Departments of Electrical Engineering, Computer Science and Engineering, and Bioengineering, Pennsylvania State University, University Park, PA 16802 USA

William E. Higgins: weh2@psu.edu

Abstract

Bronchoscopy is a major step in lung cancer staging. To perform bronchoscopy, the physician uses a procedure plan, derived from a patient's 3D computed-tomography (CT) chest scan, to navigate the bronchoscope through the lung airways. Unfortunately, physicians vary greatly in their ability to perform bronchoscopy. As a result, image-guided bronchoscopy systems, drawing upon the concept of CT-based virtual bronchoscopy (VB), have been proposed. These systems attempt to register the bronchoscope's live position within the chest to a CT-based virtual chest space. Recent methods, which register the bronchoscopic video to CT-based endoluminal airway renderings, show promise but do not enable continuous real-time guidance. We present a CT-video registration method inspired by computer-vision innovations in the fields of image alignment and image-based rendering. In particular, motivated by the Lucas-Kanade algorithm, we propose an inverse-compositional framework built around a gradient-based optimization procedure. We next propose an implementation of the framework suitable for image-guided bronchoscopy. Laboratory tests, involving both single frames and continuous video sequences, demonstrate the robustness and accuracy of the method. Benchmark timing tests indicate that the method can run continuously at 300 frames/s, well beyond the real-time bronchoscopic video rate of 30 frames/s. This compares extremely favorably to the 1 s/frame speeds of other methods and indicates the method's potential for real-time continuous registration. A human phantom study confirms the method's efficacy for real-time guidance in a controlled setting, and, hence, points the way toward the first interactive CT-video registration approach for image-guided bronchoscopy. Along this line, we demonstrate the method's efficacy in a complete guidance system by presenting a clinical study involving lung cancer patients.

Index Terms

Bronchoscopy; endoscopy; image alignment; image-based rendering; lung; lung cancer; registration; surgical guidance/navigation; virtual/augmented reality; virtual bronchoscopy; virtual endoscopy; X-ray imaging and computed tomography

I. Introduction

Lung cancer is the leading cause of cancer death in the United States [1]. The current practice for lung cancer staging involves two steps [2]–[4].

1. Procedure planning: The physician inspects a patient's three-dimensional (3D) multi-detector computed-tomography (MDCT) chest scan to select diagnostic targets (suspect nodules, lymph nodes) and to plan lung airway routes leading to the targets.
2. Bronchoscopy: The physician uses the procedure plan to navigate a bronchoscope through the airways to each target and then performs tissue biopsies.

Two difficulties, however, make this process problematic: 1) physicians poorly translate mentally derived MDCT-based procedure plans to bronchoscopy [5], [6]; 2) considerable skill variations exist between physicians who perform bronchoscopy [7]–[9]. Fluoroscopy, CT fluoroscopy, and endobronchial ultrasound assist with target confirmation and localization, but offer no guidance in navigating the bronchoscope through the lung airways [10], [11]. Also, CT-guided percutaneous biopsy is a common alternative to bronchoscopy, but it can result in a high rate of pneumothorax [2], [12].

Virtual bronchoscopy (VB) has shown considerable promise for assisting guided bronchoscopic navigation [13]–[21]. VB, a form of virtual endoscopy specific to the chest, involves graphical rendering of the airway endoluminal surfaces, as captured in a 3D MDCT scan [22], [23]. A sequence of VB views provides a 3D road map through the MDCT-derived “virtual” airways similar to the video stream provided by “real” videobronchoscopy. In fact, it is well accepted that VB views strongly resemble the structures depicted in bronchoscopic video [13]–[16], [24]. Or, stated differently, the two basic assumptions of VB are as follows: 1) the airway surfaces are locally rigid within the 3D volume directly observed by the real bronchoscope; and 2) MDCT-derived endoluminal surfaces give an accurate geometric representation of the real airway surfaces.

Drawing upon VB, researchers have recently proposed two types of bronchoscopic guidance systems: 1) electromagnetic systems; and 2) VB only systems. Electromagnetic systems link an MDCT-derived VB sequence to a sensor probe, which is tracked through the lungs during bronchoscopy [25]–[28]. Unfortunately, these systems require substantial extra hardware and set up, do not provide direct bronchoscopic vision during navigation, are susceptible to interference from metal objects, suffer from breathing-motion errors, and may produce pneumothoraces. Hybrid approaches have attempted to ameliorate motion errors, but many of the aforementioned issues still remain [29]–[31].

VB only bronchoscopy-guidance systems, on the other hand, do not require expensive sensor hardware or excessive set-up prior to bronchoscopy [6], [20], [21]. In addition, since the physician navigates the bronchoscope itself toward a target, VB only systems enable direct vision along the entire airway route. These systems, however, do not provide a direct linkage of VB information to the live bronchoscopic video. The physician must mentally “register” VB guidance views to the bronchoscopic video while navigating the bronchoscope.

On another front, motivated by research in image-guided surgery systems [32]–[34], other VB only systems have attempted to employ automatic image registration between the MDCT-based VB guidance space and real bronchoscopic video space [8], [35]–[42]. Fig. 1 summarizes the top-level registration scenario. Two distinct 3D spaces representing the patient's chest (the *World*) arise: a *virtual space* derived from the patient's 3D MDCT chest scan; and a *real space* drawn from the bronchoscope's video stream. During procedure planning, automated methods use the patient's 3D MDCT scan to define the airway tree, airway endoluminal surfaces, and guidance routes. During bronchoscopy, the guidance system draws upon this planning data to facilitate image-guided bronchoscopy.

Bricault *et al.*'s pioneering effort combined concepts from mathematical morphology, shape-from-shading analysis, and a heuristic registration approach. Their method, however, draws upon many modeling assumptions that resulted in large registration errors [35]. Helferty *et al.* attempted to maximize the similarity between the VB view and bronchoscopic video by locally searching for the optimal VB viewpoint within the 3D MDCT space [8], [36], [37]. They applied the concept of normalized mutual information (NMI), while emphasizing the information-rich dark regions (airway lumen), to give a local pose-optimization method for static single-frame registration. Unfortunately, because of their high computational demand, the two methods above are not viable for continuous registration and tracking during a live interactive procedure.

In an effort to enable live interactive bronchoscopy guidance, other approaches have proposed some form of video tracking in conjunction with CT-video registration to strive for continuous synchronization of the two spaces. Deguchi and Luo, using the observation that much of a typical bronchoscopic video frame is uninformative and hence not useful for registration, proposed a registration/tracking method that only draws upon video data residing in structurally interesting subblocks [38]–[40]. Their approach uses features, such as intensity standard deviation and saturation level, to characterize subblock information content. Focusing on the selected subblocks, the approach then employs a modified mean-squared error measure and Powell's method to find the optimal matching VB view. Thus, unlike Helferty's method, the Deguchi–Luo method attempts continuous video-sequence processing by using the results of the previous frame to initialize the analysis for the next frame.

Rai *et al.* also proposed an interleaved registration/tracking approach [41]. After an initial NMI-based CT-video registration to synchronize the real and virtual spaces, the method then processes the bulk of the incoming bronchoscopic video using real-time optical-flow (OF) analysis in conjunction with video warping to track the motion of the real bronchoscope. In particular, for each pair of successive video frames, the method determines the optimal warped version of the current video frame that best matches the subsequent incoming video frame, where the warping mechanism uses real-time OF analysis to estimate the bronchoscope's current pose. Unfortunately, optical flow tends to accumulate drift error over time. In addition, this tracking phase does not directly involve the MDCT-based virtual space. Hence, the updated pose derived from the video-only analysis gradually loses true alignment with the virtual space. To correct for these issues, the method periodically invokes single-frame NMI-based registration (e.g., every $N = 10$ video frames) to resynchronize the poses of the real and virtual spaces. As pointed out later in Section III-C, this registration results in constant intolerable several-second halts during bronchoscopy, making the method unusable for interactive live studies.

As a forerunner to their work of [41], Rai *et al.* had earlier proposed a registration/tracking approach that interleaves feature-based registration and the same optical-flow-based video tracking and warping as used in their later work [42]. Unfortunately, they found that image

features, based on the scale invariant feature transform (SIFT), did not offer sufficient robustness to function effectively over a continuously changing bronchoscopic video sequence [43], [44].

Of all of the methods in [8], [36]–[42], only the method of Helferty has been applied successfully to live patient studies [8], [37]. Nevertheless, as discussed later in Section III-C, all of these methods, including those methods that combine registration and tracking, typically require on the order of 1 s or more processing time per frame, much slower than the 30 frame/s video stream rate of standard bronchoscopes. Hence, no existing method permits continuous real-time multi-frame bronchoscopic video analysis along an extended airway route; i.e., they are not practical for interactive live guidance of patient studies.

We present a new framework that enables real-time continuous registration between the MDCT volume and bronchoscopic video. As elaborated on more below, our framework employs an image-alignment method known as the inverse-compositional formulation for extremely rapid image registration. Because the existing image-alignment methods do not provide the 3D information necessary to solve our problem, we also adapt a technique known as image-based rendering that enables MDCT-assisted 3D warping of the bronchoscopic video source. The method constitutes part of a large system under development for the planning and guidance of bronchoscopy [9], [45]–[51].

Our proposed framework for solving the CT-video registration posed by Fig. 1 is inspired by computer-vision innovations in the field of 2D-2D image alignment [44], [52]–[58]. These methods consider the problem of linking together 2D images taken from different viewpoints of a large 3D scene. A typical application involves stitching together a series of small 2D photographs to form a larger panoramic view of a scene [44], [54]. To accomplish this task, these methods parameterize the spatial relationships between the 2D images and employ some form of iterative algorithm drawing upon image warping to perform the alignment. Example methods include the pioneering Lucas–Kanade algorithm, which casts the problem as a Gauss–Newton minimization, and the computationally efficient inverse-compositional formulation, which uses derivative information that is calculated once instead of during every iteration [52], [55].

Unfortunately, these methods only give an affine transformation (eight-parameter homography) enabling alignment of the images produced by two cameras observing a 3D scene, which is precisely the situation we have for our problem of Fig. 1. They do not give the complete set of extrinsic parameters relating the two cameras; i.e., the information defining the relative 3D positions of the two cameras with respect to each other [54], [56]. Solving this problem requires a 3D-2D alignment, whereby a 2D image is aligned to a 3D scene. This, however, requires that 3D positional knowledge be known for one of the camera images, something which is generally unavailable.

Fortunately, in the scenario of bronchoscopy, this information is readily available from the 3D MDCT scan and accompanies every VB view. In our framework, the bronchoscopic video serves as the 2D image source and registrations occur on the 3D MDCT-based internal airway-tree surface model, which gives VB views and associated 3D positional knowledge. Thus, we extend the previously proposed 2D-2D image-alignment methods to handle general 3D camera motion, allowing us to iteratively solve for the extrinsic camera parameters and localize it within its environment. We then use this camera information to synthesize warped video views and perform the necessary 3D-2D alignment.

Overall, our method continuously synchronizes the real and virtual spaces for every incoming video frame, unlike approaches that interleave video tracking and periodic registration. To do this, our method performs image registration for all incoming video. The

method draws upon a 3D-to-2D image-alignment algorithm to register the video to MDCT-based VB views. Registration is facilitated by 3D-to-2D image-based rendering to warp video frames to MDCT-based VB views. This is in contrast to the Rai *et al.* approach that performs a fundamentally different OF-based video-to-video warping for bronchoscope motion tracking, which does not result in a true synchronization of the real and virtual spaces [41]. Also, our new method performs registration for every video frame at a rate far exceeding real-time operation, as opposed to the Rai *et al.* approach which has the fatal flaw of requiring guidance to halt every few seconds for a NMI-based CT-video registration. Thus, as our results later demonstrate, our method enables truly interactive image-guided bronchoscopy.

Section II discusses the theory of our framework and outlines a series of computationally efficient algorithms for continuous multi-frame registration and guided bronchoscopy. Section III validates the efficacy of the proposed framework and illustrates the utility of our complete guidance system for human lung cancer patients. Finally, Section IV offers concluding remarks.

II. Methods

The key step during image-guided bronchoscopy involves keeping the guidance system synchronized with the bronchoscope during navigation along an airway route. This requires aligning the virtual space with the real space. We do this by registering MDCT-based VB views $\mathbf{I}_{CT}(u, v)$ to bronchoscopic video frames $\mathbf{I}_V(u, v)$. Note that both of these image sources act as cameras imaging their respective 3D spaces during bronchoscopy, whereby each produces 2D endoluminal (interior) views of locations within the 3D airway tree (Fig. 1) [8], [37]. Before registration, the virtual space's misalignment with the real space is given by an unknown 6-D pose vector

$$\Theta = [\alpha \quad \beta \quad \gamma \quad t_x \quad t_y \quad t_z]^T$$

where, per the standard camera-imaging geometry, (α, β, γ) are Euler rotation angles and (t_x, t_y, t_z) specify translations along the three coordinate axes [59]. Image registration involves finding the optimal pose $\Theta = \Theta_o$ that enables alignment of the two spaces. This requires solving the optimization problem

$$\Theta_o = \arg \left\{ \min_{\Theta \in N_{\Theta_t}} \mathbf{C}(\mathbf{I}_{CT}^\Theta, \mathbf{I}_V) \right\} \quad (1)$$

where $\mathbf{I}_V(u, v)$ is the target bronchoscopic video frame, $\mathbf{I}_{CT}^\Theta(u, v)$ denotes a VB view moved an amount Θ from base view $\mathbf{I}_{CT}(u, v)$, Θ_t is an initial estimate of the desired pose, N_{Θ_t} specifies a local search neighborhood about Θ_t , and $\mathbf{C}(\cdot, \cdot)$ is a cost function. After registration, both cameras will face the same World location, implying that the coordinate systems of the two spaces have become aligned.

This section describes our approach for addressing the registration problem (1). Section II-A lays out the basic camera geometry and elements for solving (1). Section II-B provides the mathematical theory and algorithms defining our registration framework. Finally, Section II-C gives implementation details.

A. Problem Set Up

Let the virtual space be defined by a Euclidean coordinate system, whose origin and axes match those of the 3D MDCT scan. Similarly, let the real space be defined by a camera-centric Euclidean coordinate system, whose origin is at the bronchoscope tip and whose axes align with the image rows, image columns, and principal axis of the bronchoscope.

Consider a 3D point on the airway-tree endoluminal surface of the 3D MDCT scan given in homogeneous coordinates by

$$\mathbf{p} = [x \ y \ z \ 1]^T. \quad (2)$$

Under the basic assumptions of VB, there exists an unknown 3D rigid-body transformation

$$\mathbf{p}' = \mathbf{M}_\Theta \cdot \mathbf{p} \quad (3)$$

that relates virtual point $\mathbf{p}$ to a corresponding real-space point

$$\mathbf{p}' = [x' \ y' \ z' \ 1]^T \quad (4)$$

on the real airway endoluminal surface, thereby locally aligning the real and virtual surfaces as in the context of registration problem (1) when $\Theta = \Theta_o$. Equation (3) introduces the 4×4 matrix

$$\mathbf{M}_\Theta = \begin{bmatrix} \mathbf{R}(\alpha, \beta, \gamma) & \begin{matrix} t_x \\ t_y \\ t_z \end{matrix} \\ 0 & 1 \end{bmatrix} \quad (5)$$

where $\mathbf{M}_\Theta$ is a member of the special Euclidean group $SE(3)$ applicable to rigid-body transformations [60] and $\mathbf{R}(\alpha, \beta, \gamma)$ is the standard 3D 3×3 rotation matrix given in [41] and [59]. Now, 3D point $\mathbf{p}'$ maps onto camera point

$$\mathbf{c}' = [u \ v \ z' \ 1]^T \quad (6)$$

in the real bronchoscope camera's 2D image plane $\mathbf{I}_V$ via the perspective projection

$$\mathbf{c}' = \mathbf{P}\mathbf{p}' \quad (7)$$

where the camera matrix

$$\mathbf{P} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (8)$$

modeling the camera's intrinsic parameters assumes that the real camera abides by a pinhole camera model with focal length f [59]. Combining (2) and (8), the correspondence between virtual-space surface point $\mathbf{p}$ and its analogous real-space camera point $\mathbf{c}'$ is given by

$$\mathbf{c}' = \mathbf{P}\mathbf{M}_\Theta \cdot \mathbf{p}. \quad (9)$$

Or, in 2D camera-plane coordinates

$$(u, v) = \left(\frac{fx'}{z'}, \frac{fy'}{z'} \right) \quad (10)$$

per (4) and (6) and (8). Note that we deliberately preserve the depth (z) information of the endoluminal surface points imaged on the camera plane and maintain homogeneous coordinates in (6) and (10), as this will be useful later when we apply the concept of image-based rendering for transforming the video frames $\mathbf{I}_V$.

Also, it is well accepted that (9) and (10) apply to both the real- and virtual-space cameras involved in image-guided bronchoscopy [8], [35], [61]. (Section II-C points out that we apply a calibration process to match the intrinsic parameters of the real- and virtual-space cameras.) In particular, suppose that the virtual camera is placed at a pose Θ within the virtual-space airway tree. This produces a VB view $\mathbf{I}_{CT}^\Theta$ of the observed endoluminal surfaces projected onto the virtual camera's image plane via (9) and (10). Clearly, when $\Theta = \Theta_o$, a point $\mathbf{p}$ on the endoluminal surfaces projects onto the same 2D point (u, v) in both the virtual-space VB view $\mathbf{I}_{CT}^{\Theta_o}(u, v)$ and real-space video frame $\mathbf{I}_V(u, v)$, and (u, v) should therefore have similar appearance in both images; i.e.,

$$\mathbf{I}_{CT}^{\Theta_o}(u, v) \approx \mathbf{I}_V(u, v), \quad \forall (u, v). \quad (11)$$

For image-guided bronchoscopy, the pose Θ_o is unknown, and the general image-registration scenario involves trying to find the optimal pose $\Theta = \Theta_o$ to realize (11). We find Θ_o by solving the optimization problem (1).

For this work, we use the illumination-robust normalized cross-correlation cost

$$\mathbf{C}(\mathbf{I}_{CT}^\Theta, \mathbf{I}_V) = - \sum_{u,v} [\bar{\mathbf{I}}_{CT}^\Theta(u, v) \cdot \bar{\mathbf{I}}_V(u, v)] \quad (12)$$

successfully used by Mori *et al.*, where

$$\bar{\mathbf{I}}_{CT}(u, v) = \frac{\mathbf{I}_{CT}(u, v) - \mu_{CT}}{\sigma_{CT}}$$

and

$$\bar{\mathbf{I}}_V(u, v) = \frac{\mathbf{I}_V(u, v) - \mu_V}{\sigma_V}$$

are normalized forms of $\mathbf{I}_{CT}$ and $\mathbf{I}_V$, and $(\mu_{CT}, \sigma_{CT}^2)$ and (μ_V, σ_V^2) denote the sample means and variances of the pixels in images $\mathbf{I}_{CT}$ and $\mathbf{I}_V$, respectively [38]. Optimization (1) becomes simpler when the cost is posed in a least-squares normalized sum-of-squared differences (NSSD) form

$$\mathbf{C}_{NSSD}(\mathbf{I}_{CT}^\Theta, \mathbf{I}_V) = \sum_{u,v} [\bar{\mathbf{I}}_{CT}^\Theta(u, v) - \bar{\mathbf{I}}_V(u, v)]^2. \quad (13)$$

By expanding the right-hand side of (13) as

$$\mathbf{C}_{NSSD}(\mathbf{I}_{CT}^\Theta, \mathbf{I}_V) = \sum_{u,v} \left[\left(\bar{\mathbf{I}}_{CT}^\Theta(u, v) \right)^2 - 2 \cdot \left(\bar{\mathbf{I}}_{CT}^\Theta(u, v) \cdot \bar{\mathbf{I}}_V(u, v) \right) + \left(\bar{\mathbf{I}}_V(u, v) \right)^2 \right] \quad (14)$$

it is easy to show that (12) and (13) are affinely related, and, hence, give equivalent optima. To see this, first note that the first and last terms of (14), which have sample means equal to zero and sample variances equal to one, individually sum to a constant. This leaves the middle term, which is obviously proportional to the right-hand side of (12). Therefore, given the equivalence of (12) and (13), we can instead use cost function (13) in (1) to give the following nonlinear least-squares optimization problem:

$$\Theta_o = \arg \left\{ \min_{\Theta \in N_{\Theta_t}} \sum_{u,v} [\bar{\mathbf{I}}_{CT}^\Theta(u, v) - \bar{\mathbf{I}}_V(u, v)]^2 \right\}. \quad (15)$$

B. Inverse-Compositional Formulation via Image-Based Rendering

The obvious approach to solving (15) involves incrementally moving the virtual camera toward the real camera's unknown position. This leads to the Lucas–Kanade-motivated forward-compositional formulation [52], [56]. Unfortunately, the forward-compositional formulation has a well-known high computational complexity [56]. In addition, in our scenario, this method incurs a considerable extra burden, since it requires the calculation of a very large number of new VB views $\mathbf{I}_{CT}^\Theta$ during optimization. Nevertheless, because this formulation is a viable method for solving (15) and because Section III considers its performance, the Appendix summarizes it.

As an alternative, we propose a far more computationally efficient inverse-compositional formulation that is largely equivalent to the forward-compositional approach [53], [54]. Instead of moving the virtual camera towards the real camera's unknown pose, we can derive the same relative rigid transformation by instead moving the real camera toward the virtual camera. The real camera being the bronchoscope is, of course, fixed during optimization, as we have no control over its movement. Thus, we must approximate the real camera's movement and in turn the updated video frame during optimization.

Fortunately, this process is possible through a technique referred to as image-based rendering [44], [58]. Image-based rendering has been widely considered by the computer-vision community for synthesizing photographs and video shots that appear to be captured

from alternate viewpoints. This synthesis requires 3D positional knowledge of each image point (u, v) .

For our scenario (15), we can synthesize warped versions of at alternate viewpoints by gleaming the necessary 3D information from the MDCT-based virtual space. In particular, as with (6) and (7), a 3D surface point maps onto a virtual-space camera point via $\mathbf{c}$ or $\mathbf{c} = \mathbf{P}\mathbf{p}$, analogous to (10), $\mathbf{p}$ maps onto

$$(u, v) = \left(\frac{fx}{z}, \frac{fy}{z} \right) \quad (16)$$

in virtual-camera VB view $\mathbf{I}_{CT}$. Note that each image point of (u, v) arises from a uniquely observed endoluminal surface point $\mathbf{p}$ and has a known depth z associated with it. Thus, we can perform an inverse projection of image point (u, v) to find the unique 3D airway-surface point $\mathbf{p}$ that maps to (u, v) via

$$\mathbf{Q}(u, v) = \begin{bmatrix} x & y & z & 1 \end{bmatrix}^T \quad (17)$$

where $\mathbf{Q}(u, v)$ signifies the inverse projection function. Note in passing that $\mathbf{P}\mathbf{Q}(u, v) = \mathbf{c}$. Also, note that the information for (17) is automatically available from the depth map created during the rendering of VB view $\mathbf{I}_{CT}$. For the final step, we then map the obtained point $\mathbf{Q}(u, v)$ into the bronchoscope's 2D image plane $\mathbf{I}_V$ using (9)

$$\mathbf{c}' = \mathbf{P}\mathbf{M}_\Theta \mathbf{Q}(u, v) \quad (18)$$

or in camera coordinates analogous to (10)

$$(u, v) = \mathbf{W}(u, v; \Theta). \quad (19)$$

Using the parlance of image alignment, we will refer to function $\mathbf{W}(\cdot, \cdot; \Theta)$ as a warping function [56]. Thus, in (19), $\mathbf{W}(\cdot, \cdot; \Theta)$ warps a 2D virtual camera point (u, v) into a real camera point $\mathbf{W}(u, v; \Theta)$ by combining (17) and (18) and (10) to give a view

$$\mathbf{I}_V^\Theta(u, v) = \mathbf{I}_V(\mathbf{W}(u, v; \Theta)).$$

Given the above, it becomes possible to swap the roles of the real and virtual spaces in the optimization. This role reversal effectively treats the VB view as a static image, allowing it to be rendered only once during registration. We assume without loss of generality that $\mathbf{I}_V$ is at an unknown initial pose Θ_i , while VB view $\mathbf{I}_{CT}$ is at known fixed pose $\Theta = \mathbf{0}$.

The inverse compositional method switches the roles of $\mathbf{I}_{CT}$ and $\mathbf{I}_V$ by warping $\mathbf{I}_V$ so that it approximately matches a fixed target $\mathbf{I}_{CT}$. In particular, to solve (15), the inverse-compositional formulation minimizes

$$\hat{\mathbf{C}}(\mathbf{I}_{CT}, \mathbf{I}_V, \Theta) = \sum_{u, v} [\bar{\mathbf{I}}_{CT}^{\Delta\Theta}(u, v) - \bar{\mathbf{I}}_V^\Theta(u, v)]^2 \quad (20)$$

with respect to $\Delta\Theta$, where Θ represents the pose difference between the real and virtual spaces.

To solve (15) using (20), we assume that the pose difference $\Theta = \mathbf{0}$ and iteratively solve for the actual Θ using the update rule

$$\mathbf{M}_\Theta \leftarrow \mathbf{M}_\Theta \cdot \mathbf{M}_{\Delta\Theta}^{-1}. \quad (21)$$

Paralleling the Lucas–Kanade formalism, we consider an iterative Gauss–Newton gradient-descent algorithm for the optimization [56]. To derive the necessary algorithm elements, we first linearize (20) by approximating $\bar{\mathbf{I}}_{\text{CT}}^{\Delta\Theta}(u, v)$ by its first-order Taylor-series expansion about $\Delta\Theta = \mathbf{0}$

$$\hat{\mathbf{C}}(\mathbf{I}_{\text{CT}}, \mathbf{I}_V, \Theta) \approx \sum_{u,v} [\bar{\mathbf{I}}_{\text{CT}}(u, v) + (\nabla \bar{\mathbf{I}}_{\text{CT}} \Delta\Theta)_{\Theta=\mathbf{0}} - \bar{\mathbf{I}}_V^\Theta(u, v)]^2 \quad (22)$$

where

$$\nabla \bar{\mathbf{I}}_{\text{CT}} \equiv \left[\frac{\partial \bar{\mathbf{I}}_{\text{CT}}}{\partial \Theta} \right] = \begin{bmatrix} \frac{\partial \bar{\mathbf{I}}_{\text{CT}}}{\partial \alpha} & \frac{\partial \bar{\mathbf{I}}_{\text{CT}}}{\partial \beta} & \frac{\partial \bar{\mathbf{I}}_{\text{CT}}}{\partial \gamma} & \frac{\partial \bar{\mathbf{I}}_{\text{CT}}}{\partial t_x} & \frac{\partial \bar{\mathbf{I}}_{\text{CT}}}{\partial t_y} & \frac{\partial \bar{\mathbf{I}}_{\text{CT}}}{\partial t_z} \end{bmatrix} \quad (23)$$

is the gradient of $\bar{\mathbf{I}}_{\text{CT}}$ with respect to Θ .

Now, differentiating (22) with respect to $\Delta\Theta$ and setting to 0 gives the optimal Gauss–Newton descent direction

$$\Delta\Theta = \mathbf{H}^{-1} \cdot \sum_{u,v} [\nabla \bar{\mathbf{I}}_{\text{CT}}]^T \cdot [\bar{\mathbf{I}}_V^\Theta(u, v) - \bar{\mathbf{I}}_{\text{CT}}(u, v)] \quad (24)$$

where

$$\mathbf{H} = \sum_{u,v} [\nabla \bar{\mathbf{I}}_{\text{CT}}]^T \cdot [\nabla \bar{\mathbf{I}}_{\text{CT}}] \quad (25)$$

is the Gauss–Newton approximation to the 6×6 Hessian matrix of $\bar{\mathbf{I}}_{\text{CT}}$.

Drawing upon this development, the inverse-compositional formulation addresses problem (15) by registering video frame $\mathbf{I}_V$ at an unknown pose to VB view $\mathbf{I}_{\text{CT}}$ at a known pose; i.e., it moves the real-space camera toward the fixed virtual-space camera. In particular, the method iterates pose-update equations, (21) and (24), until (20) converges. As a major observation, note that the quantities $\{\bar{\mathbf{I}}_{\text{CT}}, \nabla \bar{\mathbf{I}}_{\text{CT}} \mathbf{H}^{-1}\}$ remain constant during iteration of (24). Thus, as we will see in Section II-C, the inverse-compositional formulation achieves considerable computational efficiency by having these quantities precomputed. Upon convergence, Θ_o gives the optimal pose aligning the real and virtual spaces and

$$\mathbf{I}_{\text{CT}}^{\Theta_o} = \mathbf{I}_{\text{CT}} \left(\mathbf{W}^{-1}(u, v; \Theta_o) \right) \quad (26)$$

gives the VB view best matching video frame $\mathbf{I}_V$, where $\mathbf{W}(u, v; \Theta_o)$ represents the optimal warp.

Algorithms 1–3 incorporate the method into a system-level strategy for the two-step lung cancer staging process. As part of procedure planning, Algorithm 1 performs important precomputation, while, during bronchoscopy, Algorithms 2–3 enable continuous registration.

During procedure planning, we first derive an airway-tree model from the patient's 3D MDCT chest scan. The airway-tree model consists of the airway endoluminal surfaces, associated surface depth maps, airway centerlines, and guidance routes. We use the heavily validated automated methods of [8], [45]–[48] to derive this model. Because bronchoscopy occurs inside the airways, we know that the airway tree constitutes the working navigation environment and, hence, encompasses the possible poses that the bronchoscope could assume during bronchoscopy. Given the airway-tree model, Algorithm 1 precomputes virtual-space quantities $\{\bar{\mathbf{I}}_{CT}, \nabla \bar{\mathbf{I}}_{CT} \mathbf{H}^{-1}\}$ used by update (24) over a reference pose set Θ_{ref} , where Θ_{ref} spans the interior of the 3D airway tree. In particular, Θ_{ref} consists of a set of reference poses Θ_i that finely span individual viewing sites along the airway centerlines.

Algorithm 1

Data precomputation for reference pose set Θ_{ref}

```

1: given: MDCT-based airway-tree model computed during Procedure Planning
2: for all reference poses  $\Theta_i \in \Theta_{\text{ref}} = \{\Theta_1, \Theta_2, \dots, \Theta_{N_{\text{ref}}}\}$  do
3:
   Calculate VB view  $\bar{\mathbf{I}}_{CT}^{\Theta_i}(u, v)$  and depth map  $d^{\Theta_i}(u, v) = z$ 
4:
   Compute component images  $\left\{ \frac{\partial \bar{\mathbf{I}}_{CT}^{\Theta_i}}{\partial \alpha}, \frac{\partial \bar{\mathbf{I}}_{CT}^{\Theta_i}}{\partial \beta}, \dots, \frac{\partial \bar{\mathbf{I}}_{CT}^{\Theta_i}}{\partial t_z} \right\}$  for  $\nabla \bar{\mathbf{I}}_{CT}^{\Theta_i}$ 
5:
   Compute inverse Hessian  $\mathbf{H}_{\Theta_i}^{-1}$ 
6:
return  $\left\{ \bar{\mathbf{I}}_{CT}^{\Theta_i}(u, v), d^{\Theta_i}(u, v), \nabla \bar{\mathbf{I}}_{CT}^{\Theta_i}, \mathbf{H}_{\Theta_i}^{-1} \right\}$  for all poses  $\Theta_i \in \Theta_{\text{ref}}$ 

```

Algorithm 2

Continuous registration over a video sequence $\mathbf{I}_V(k)$, $k = 0, 1, 2, \dots$

```

1:
   Present first target VB view  $\mathbf{I}_{CT}^{\Theta_t}$  at the guidance route's beginning
2:
   Navigate bronchoscope close to  $\mathbf{I}_{CT}^{\Theta_t}$  so that real space roughly aligns with virtual space
3: for each incoming bronchoscopic video frame  $\mathbf{I}_V(k)$ ,  $k = 0, 1, 2, \dots$  do
4:
    $\mathbf{I}_{CT}^{\Theta_o} \leftarrow \text{Algorithm 3} \left( \mathbf{I}_V(k), \mathbf{I}_{CT}^{\Theta_t} \right)$ 
5:
   Output  $\mathbf{I}_{CT}^{\Theta_o}(u, v)$ 
6:
    $\mathbf{I}_{CT}^{\Theta_t} = \mathbf{I}_{CT}^{\Theta_o}(u, v)$ 

```

During bronchoscopy, Algorithm 2 lays out the image-guided bronchoscopy process. It begins by having the guidance computer present an initial VB view at the beginning of a guidance route (line 1), generally near the main carina. The bronchoscopist then inserts the bronchoscope into the patient's chest and maneuvers the device close to the given VB view (line 2). At this point, Algorithm 2 invokes continuous video capture and registration (lines 3–6). The bronchoscopist now freely moves the bronchoscope along the guidance route, with registrations continuously occurring at real-time video frame rates. In particular, for each new video frame, Algorithm 2 calls Algorithm 3 at line 4 to invoke a registration of the real and virtual spaces.

Algorithm 3

Inverse-Compositional Registration.

```

1:  given:
2:      1) video frame  $I_V$  at unknown pose
3:      2) VB view  $I_{CT}$  at target global pose  $\Theta_t$ 
4:  repeat
5:      Find reference pose  $\Theta_i \in \Theta_{\text{ref}}$  closest to target pose  $\Theta_t$ 
6:      Find warp modeling relative pose between reference pose  $\Theta_i$  and target pose  $\Theta_t$  via:  $\mathbf{M}_\Theta \leftarrow \mathbf{M}_{\Theta_i}^{-1} \cdot \mathbf{M}_{\Theta_t}$ 
7:      Compute normalized warped video frame  $\bar{\mathbf{I}}_V^\Theta(u, v)$ 
8:      Compute update  $\Delta\Theta$  per (24) using  $\bar{\mathbf{I}}_V^\Theta(u, v)$  and  $\{\bar{\mathbf{I}}_{CT}^{\Theta_i}, \nabla \bar{\mathbf{I}}_{CT}^{\Theta_i}, \mathbf{H}_{\Theta_i}^{-1}\}$ 
9:      Update warp for pose difference using (21)
10:      $\mathbf{M}_{\Theta_r} \leftarrow \mathbf{M}_{\Theta_i} \cdot \mathbf{M}_\Theta$ 
11:  until cost (20) converges
12:  return  $\bar{\mathbf{I}}_{CT}^\Theta$  // Equals optimal VB view  $\bar{\mathbf{I}}_{CT}^{\Theta_o}$  per (26)

```

Algorithm 3 begins with a coarse alignment and then iteratively arrives at a precise pose. More specifically, line 5 enables the coarse alignment by allowing multiple reference poses Θ_i to be used during registration. This permits estimation of larger pose differences that could exist initially between the given video frame and target VB view; i.e., large relative to the fine granularity of the precomputed poses $\Theta_i \in \Theta_{\text{ref}}$. Generally, after a few iterations, Θ_i no longer changes at line 5. In particular, suppose at iteration $(j + 1)$, Θ_i does not change from iteration j . Then, the result of line 10 on iteration j combined with the result of line 6 during the next iteration $(j + 1)$ gives

$$\begin{aligned}
 \mathbf{M}_\Theta &= \mathbf{M}_{\Theta_i}^{-1} \cdot \mathbf{M}_{\Theta_t} \\
 &= \mathbf{M}_{\Theta_i}^{-1} \cdot \mathbf{M}_{\Theta_i} \cdot \mathbf{M}_\Theta \\
 &= \mathbf{M}_\Theta
 \end{aligned}$$

and the algorithm settles into iterating around the same global reference pose Θ_i for iterations j ; i.e., subsequent updates become incremental relative to the spacing of the reference poses and follow the mathematics of (21) and (24).

Without question, the greatest benefit of the inverse-compositional formulation is its well-known computational efficiency [56]. In fact, for our CT-video registration scenario, the inverse method as embodied by Algorithm 3 offers vastly greater computational efficiency over the forward method. This is because the forward method requires the computation of 13 VB views per *iteration*, while the inverse method only computes one VB view per *registration* (see Appendix). Our results in Section III-C clearly bear out this efficiency.

We point out that while the inverse and forward approaches result in equivalent optima for 2D and 3D data, there is no such proof of equivalence for our so-called 2.5D situation, as first noted in [62]. This is because the functions relating the pixels of the real and virtual images must form a group. Along this line, while rigid transform $\mathbf{M}_\Theta \in SE(3)$, the inclusion of $\mathbf{Q}$ and $\mathbf{P}$, which involve projecting 3D data into 2D image planes, implies that the family of warping functions $\mathbf{W}(u, v; \Theta)$ formed by composition (18) do not form a group. In particular, at the optimal pose $\Theta = \Theta_o$, the gradients of the warping functions used by the inverse and forward approaches must be equivalent for all points $\mathbf{p}$ in 3D space. Unfortunately, the optimality condition (11) only holds for points $\mathbf{p}$ on the 2D airway surface.

Nevertheless, the inverse method's use of image-based rendering for our scenario effectively synthesizes a rigid-body 3D motion of the real-space video camera. Also, this process is assisted by Algorithm 3's coarse-to-fine alignment process and by our use of a very dense reference-pose set, as discussed below. This enables a decrease in gradient errors as the overall camera pose error decreases. (Romdhani and Vetter employed a related approach for fitting 3D morphable models to 2D images of faces [63].) Practically speaking, this means that the inverse method's update applied to the synthesized real camera approaches the relative transformation between the real- and virtual-space cameras achieved by the forward approach. The empirical equivalence of the two approaches is clearly borne out in Section III.

C. Implementation Details

In our implementation of Algorithm 1, the reference poses $\Theta_i \in \Theta_{\text{ref}}$ are situated near points sampled ≈ 0.5 mm apart along the automatically computed airway centerlines. To account for the bronchoscope's possible angular deflection, we actually precompute data

$\{\bar{\mathbf{I}}_{\text{CT}}^{\Theta_i}, \nabla \bar{\mathbf{I}}_{\text{CT}}^{\Theta_i}, \mathbf{H}_{\Theta_i}^{-1}\}$ for 13 distinct Θ_i corresponding to each sampled 3-D centerline position (t_x, t_y, t_z) : one given by the centerline point and twelve others accounting for individual small $\pm\Delta$ perturbations of each parameter constituting Θ_i . Note that the bronchoscope always looks forward in the airway tree and its maximum angular deflection about an airway's central axis is generally $< 30^\circ$. In addition, we compute extra poses for points at airway bifurcations: one directed toward the proximal end of the parent branch and others directed toward the distal ends of each child branch. This accounts for the bronchoscope's greater maneuverability near bifurcations. Thus, the reference-pose set Θ_{ref} gives a very dense sampling of reference poses.

Each image constituting one of the six components of ∇_{CT} in (23) is readily approximated as a simple first-difference of perturbed VB views

$$\left[\frac{\partial \bar{\mathbf{I}}_{\text{CT}}}{\partial \Theta}(u, v) \right]_j \approx \frac{\bar{\mathbf{I}}_{\text{CT}}^{\theta_j^+}(u, v) - \bar{\mathbf{I}}_{\text{CT}}^{\theta_j^-}(u, v)}{\theta_j^+ - \theta_j^-} \quad (27)$$

where θ_j refers to a parameter $\alpha, \beta, \dots, t_z$, in pose Θ_i (Fig. 2). Next, note that instead of the Gauss–Newton Hessian approximation (25), researchers in the field of image alignment often approximate the 6×6 Hessian as an identity matrix $c\mathbf{I}$. This results in a steepest-descent optimization process for the inverse-compositional formulation [56]. While this is a viable alternative, we instead draw upon a diagonal matrix of the form

$$\mathbf{H}^{-1} = \text{diag}\{h_\alpha, h_\beta, h_\gamma, h_{t_x}, h_{t_y}, h_{t_z}\} \quad (28)$$

where $\text{diag}(\cdot)$ signifies a diagonal matrix. Equation (28) offers greater precision and faster convergence over an identity matrix [56], [64]. It also permits tunable parameter weighting, which is useful given that the pose rotation parameters typically vary more than the translational parameters (Section III). Finally, (28) avoids matrix calculation/inversion, which substantially reduces computation [56].

Our implementation of Algorithm 1 uses three methods to reduce the precomputation and concomitant data storage. First, note that the typical high-resolution 3D MDCT scans used for image-guided bronchoscopy have typical voxel resolution on the order of 0.5 mm. On the other hand, a bronchoscope’s diameter d is in the range 2.8 mm $\leq d \leq$ 6.0 mm. Thus, the automatically derived centerlines span far more airways than are actually accessible to the bronchoscope [45]. Therefore, we do not consider candidate reference poses Θ_i occurring at airway locations inaccessible to the bronchoscope; i.e., locations smaller than the bronchoscope’s diameter. This typically eliminates nearly half of the candidate poses.

As motivation for the second reduction, we exploit the observation, noted by Bricault, Deguchi, and others, that most of a typical frame is uninformative [8], [35], [38]–[40]. For registration, note that only the video-frame pixels depicting airway bifurcations and regions immediately adjacent to the bifurcations provide useful registration information. Conversely, image pixels depicting bland uninteresting wall regions “dilute” registration efficacy. For this reason, for each reference pose Θ_i , we only precompute VB-view pixels situated inside a bounding rectangle that spans the segmented airways plus a small margin. This idea can eliminate nearly 75% of the precomputation for a given Θ_i . Lastly, note that a typical video frame $\mathbf{I}_V$ depicts endoluminal airway structure having a real width ~ 20 mm. Yet, $\mathbf{I}_V$ may contain, for example, 264×264 pixels (see Table I in Section III). This corresponds to a spatial resolution (< 0.1 mm) far higher than the MDCT data (0.5 mm). Hence, we precompute VB views having on the order of 150×150 pixels. Later, during bronchoscopy, we also downsample the video frames by a factor of 2 or 4 using a multi-resolution pyramid.

Using these reductions, we observed over five human cases that the size of the precomputed data generated by Algorithm 1 ranged from a minimum of 268 MB to a maximum of 1 GB (the number of airway branches in a case ranged between 81 and 233). Precomputation time on a mid-range PC equipped with a standard video card (~ 500 MB video memory) is generally < 2 min. As VB-view generation dominates the computational load, the computation time depends almost entirely on video-hardware speed. Note that all procedure planning operations, including airway-tree model computation and physician interaction, typically takes ~ 20 min for a human patient [45]. Thus, the computation of Algorithm 1 is not disruptive to the standard clinical work flow.

Regarding Algorithm 3, in line 5 the pose Θ_i giving smallest Euclidean distance $\|\Theta_i - \Theta_t\|$ is selected as the one closest to the current target pose Θ_t . If target Θ_t is near a bifurcation, we consider all reference poses Θ_i , including the additional poses precomputed for the parent and child airway branches, and designate the closest pose as the one at a viewing direction forming the smallest angle with Θ_t ’s viewing direction—this notion recognizes the

importance of direction in reaching correct registrations while also noting the dense spacing of poses in reference set Θ_{ref} .

In line 7 of Algorithm 3, image-based rendering occurs very quickly, since it only depends on the size of v . In particular, we solve (9) and (10) for each required real-space point by mapping the 3D surface coordinates $\mathbf{Q}(u, v)$ of the precomputed $\mathcal{C}_T$ into a corresponding 2D real-space location using a relative warp $\mathbf{W}(u, v; \Theta)$ derived via line 6 and (17) and (19), and then interpolating the corresponding point (u, v) for v (Fig. 3). Doing this results in a warped image $\bar{\mathbf{I}}_v^\Theta(u, v)$ simulating what would be observed by the real bronchoscope camera if placed at the viewpoint of the virtual camera. Again, the dense spacing of poses in Θ_{ref} coupled with the high resolution of v keep warping artifacts very small.

The update computation at line 8 of Algorithm 3 only uses pixels available in both the VB view and warped video frame. Thus, the sum in update (24) excludes pixel locations (u, v) situated in the undefined (black) regions of $\bar{\mathbf{I}}_v^\Theta(u, v)$, as illustrated in Fig. 3(f). The sum also only considers pixels available in the VB view's bounding rectangle. In addition, we iterate Algorithm 3 until the cost $\hat{\mathbf{C}}$ in (20) changes ε between successive iterations. Finally, it is imperative that the fields of view of both the real and virtual cameras match per (11). Otherwise, the many pixel-to-pixel calculations made by Algorithm 3 between a VB view and a video frame cannot be computed correctly. Hence, prior to bronchoscopy, we perform an offline camera calibration that allows the software-defined virtual camera to have identical intrinsic parameters to the bronchoscope's camera; i.e., the focal length f of (8) is matched for both cameras [8], [36], [65].

III. Results

We tested the proposed methods over a wide range of situations with increasing complexity, beginning with controlled single-frame tests and culminating with live guided human studies. We summarize the presented results below.

1. Section III-B: Single-frame tests.
 - Measure registration robustness for a single frame.
 - Perform both virtual-to-virtual and virtual-to-real tests.
2. Section III-C: Tests using video sequences.
 - Consider registration performance over a complete continuous video sequence.
 - Perform both virtual-to-virtual and virtual-to-real tests.
 - Measure computation time to assess feasibility for real-time operation.
3. Section III-D: Phantom studies.
 - Consider method efficacy during live bronchoscopy.
 - Run tests in a controlled environment, free of events occurring during a live patient procedure.
4. Section III-E: Live human studies.
 - Test the methodology when integrated into a complete image-guided system.
 - Apply the system to human lung cancer patients in a real clinical scenario.

The virtual-to-virtual tests of Sections III-B and III-C measure performance independent of the image-modeling differences existing between the virtual VB views $\mathbf{I}_{CT}(u, v)$ and the bronchoscopic video frames $\mathbf{I}_V(u, v)$. (As discussed further below, we substitute VB views in place of the video frames for these tests.) The virtual-to-virtual tests also enable ground-truth-based assessment of method performance. The subsequent virtual-to-real tests in Sections III-B and III-C then give a sense of performance degradation when true video is employed for registration. Next, the video-sequence tests of Section III-C show that continuous processing in real-time is possible for the inverse-compositional formulation, thereby pointing to the method's potential for guiding live bronchoscopy. Finally, Sections III-D and III-E illustrate method utility during live bronchoscopy. Before proceeding, we first introduce the primary data sets and error metrics employed in the tests of Section III-B to III-D.

A. Test Data and Error Metrics

The tests of Sections III-B–III-D drew upon data derived from five lung cancer patients, as summarized in Table I. Three different bronchoscopes were used to generate the video data. The parenthetical numbers associated with the video-frame sizes indicate the downsample factors employed for registration. For example, for patient A, we downsampled the original 592×440 bronchoscopic video frames by a factor of 4, resulting in 148×110 frames being used for registration.

We quantified registration performance as follows. With reference to optimization problem (1), each single-frame optimization trial began with an input VB view $\mathbf{I}_{CT}^{\Theta_t}$ at starting pose Θ_t and a target video frame $\mathbf{I}_R$ at a known real-space ground-truth pose

$$\Theta_G = [\alpha^G \quad \beta^G \quad \gamma^G \quad t_x^G \quad t_y^G \quad t_z^G]^T.$$

Upon convergence, a trial gave a final output VB view $\mathbf{I}_{CT}^{\Theta_o}$ at final virtual-space optimal pose

$$\Theta_o = [\alpha^o \quad \beta^o \quad \gamma^o \quad t_x^o \quad t_y^o \quad t_z^o]^T$$

with $\Theta_o \approx \Theta_G$ if a registration method is successful. Let

$$\mathbf{p}_G = [t_x^G \quad t_y^G \quad t_z^G]^T, \quad \mathbf{p}_o = [t_x^o \quad t_y^o \quad t_z^o]^T$$

represent the 3D positions and

$$\mathbf{v}_G(\alpha^G, \beta^G, \gamma^G), \quad \mathbf{v}_o(\alpha^o, \beta^o, \gamma^o)$$

specify the direction vectors for Θ_G and Θ_o . Given these definitions, we computed the following error metrics.

a. Position error

$$e_p = \|\mathbf{p}_o - \mathbf{p}_G\|$$

which gives the Euclidean distance between the optimized and ground-truth positions.

b. Direction error

$$e_d = \cos^{-1}(\mathbf{v}_o \cdot \mathbf{v}_G)$$

defined as the angle between the direction vectors of Θ_o and Θ_G .

c. Roll error

$$e_r = \cos^{-1}(\mathbf{u}_G \cdot \mathbf{u}_o^{e_d})$$

defined as the angle required to align the up vectors, $\mathbf{u}_o$ and $\mathbf{u}_G$, of Θ_o and Θ_G after the direction vectors of Θ_o and Θ_G have been aligned [66].

d. Needle error

$$e_n = \|(\mathbf{p}_o + d_n \mathbf{v}_o) - (\mathbf{p}_G + d_n \mathbf{v}_G)\|$$

which equals the distance between two points located a distance d_n along the viewing directions of Θ_o and Θ_G .

Measuring position and direction errors separately is vital for understanding the efficacy of bronchoscopy navigation, as they separate out the effect of image scaling. Roll error e_r is important, because bronchoscopy navigation almost continuously involves device-roll (twist) maneuvers. To compute e_r , we first rotate $\mathbf{v}_o$ by angle e_d in the plane defined by $\mathbf{v}_o$ and $\mathbf{v}_G$ (provided $\mathbf{v}_o \neq \mathbf{v}_G$); this aligns the direction vectors, with $\mathbf{u}_o^{e_d}$ denoting the up vector associated with the rotated $\mathbf{v}_o$. The needle error e_n models the measurement of the position difference of a biopsy needle extending from the bronchoscope tip [8]. This is important, because diagnostic targets often occur at extraluminal sites, sites that the bronchoscope cannot completely reach. Hence, for extraluminal targets, position and direction errors could conceivably become magnified when biopsies are attempted.

B. Single-Frame Registration

The goal of these tests was to determine how well registration performs for a single static video frame. These tests are vital, because if single-frame performance is poor, then registration would be unlikely to sustain success over a continuous video sequence. We present a set of controlled studies involving a variety of representative video frames and associated ground truth. To assess single-frame registration performance, we compared three registration algorithms: the inverse-compositional (IC) method of Algorithm 3; the forward-compositional (FC) method of the Appendix; and a previously validated method drawing upon normalized mutual information (NMI) [8], [37].

For single-frame registration, we iterated the IC and FC methods until cost $\hat{\mathbf{C}}$ changed $\varepsilon = 5 \times 10^{-5}$ between successive iterations ($\hat{\mathbf{C}}$ given by (20) for IC, (30) for FC). For both the IC and FC methods, we used the following values for $\mathbf{H}^{-1}$ in (28): $h_\alpha = 50$, $h_\beta = h_\gamma = 21$, $h_{t_x} = h_{t_y} = 1.0$, $h_{t_z} = 1.7$. Because the reference poses $\Theta_i \in \Theta_{\text{ref}}$ are computed over a spacing $<$

0.5 mm, these values encourage quicker alignment of the rotational components of Θ_o , which tend to change a much greater amount. The weights are also in line with the pose changes observed experimentally (see results below).

We used data from three patients to perform the single-frame tests. The ground-truth test set consisted of 10 video frames from patient A and 7 frames each from patients B and C. These frames captured a wide range of typical pose and airway-morphology variations and also considered video drawn from three different bronchoscopes. To establish a video frame's ground-truth pose Θ_G , we employed manual registration to derive the best-matching VB view (Fig. 4). For the single-frame tests, we considered two distinct subsets of tests.

1. Virtual-to-virtual: These tests involved registering the virtual space to a target VB view acting as the “video” frame $\mathbf{I}_V$. By removing the image-modeling differences between the two sources, these tests have two distinct advantages.
 - a. They admit the possibility of “perfect” registration, since both sources draw views from the same MDCT scan data.
 - b. They separate out the effect of source differences and focus strictly on the performance of the registration method.
2. Virtual-to-real: These tests involved registering the virtual space to a target ground-truth video frame $\mathbf{I}_V$. Regarding these tests.
 - a. They admit an added level of complexity over the virtual-to-virtual tests to the overall registration problem. Hence, they enable us to measure the impact of source differences.
 - b. They represent the situation encountered during “real” bronchoscopy.

Given this introduction, the first set of tests measured the sensitivity of registration to parameter variations in the starting pose Θ_t . In particular, the tests measured how much Θ_t could be perturbed one parameter at a time independently with respect to Θ_G , while still enabling 100% successful registration over all test frames. For each parameter, we made successively larger perturbations and then ran registration. We continued this process for both positive and negative perturbations until registration failed. We deemed a particular $\mathbf{I}_{CT}^{\Theta_o}$ to be a registration failure if it reached any of the following criteria:

$$e_p > 5.0 \text{ mm}, \quad e_d > 10^\circ, \quad e_r > 15^\circ. \quad (29)$$

Note that criteria (29) represent *very conservative* bounds for bronchoscopy. In particular, as the follow-up continuous-sequence results of Section III-C clearly demonstrate, they are much more stringent than necessary for successful continuous multi-frame registration. Nevertheless, they offer a useful benchmark for assessment of single-frame performance.

Table II lists the perturbation ranges that enabled 100% successful registration over all test frames, while Table III focuses on the results for patient A. These results indicate that the more computationally efficient inverse-compositional formulation performs comparably to the forward-compositional method, and both methods offer substantially better robustness to pose variations than NMI-based registration. Interestingly, both IC and FC appear to be very robust to angular variations. As expected, the methods exhibit greater robustness for the virtual-to-virtual tests. Yet, the performance degradation when using real video was not large. For example, for the patient A tests (Table III), source differences between $\mathbf{I}_{CT}$ and $\mathbf{I}_V$ only resulted in a 20% decrease in the perturbation ranges. This attests to the veracity of the image model employed for the MDCT-based VB views [37].

We point out that the ranges in Tables II and III tend to be very conservative—for every test frame, registration was frequently successful for much larger parameter deviations. For instance, for virtual-to-real registration, the methods typically tolerated α/β (pitch/yaw) variations near 25° and γ (roll) variations near 60° . Also, registration often succeeded for starting poses > 10 mm from the target location. Note that 10 mm is a very large positional deviation, given that the segmental/lobar airways typically have lengths between 5–30 mm and diameters between 2–10 mm.

Additionally, note that the position change $\Theta \mathbf{p} = [\Theta t_x \ \Theta t_y \ \Theta t_z]^T$ represented between successive video frames, $\mathbf{I}_V(k)$ and $\mathbf{I}_V(k+1)$, during bronchoscopy tends to be small. For example, a bronchoscope velocity of 10 mm/s translates to $\|\Delta \mathbf{p}\| = 0.33$ mm (video frame rate = 30 frames/s, a positional change well under the ranges listed in Tables II and III). Hence, registration is likely to be successful for all practical bronchoscope navigation velocities. On the other hand, we caution that for a video frame not in the test data set, registration failure can occur for perturbations smaller than the ranges indicated in the tables.

For the next set of tests, we measured registration accuracy and success rate for the situation where we subjected all starting-pose parameters to random simultaneous perturbations. This test employed the patient A video frames and considered 20 randomly perturbed poses for each frame. Each starting pose Θ_t was generated through the following sequence of perturbations for a given Θ_G : 1) translation performed in the $\mathbf{v}_G$ plane with radius distributed uniformly on [0 mm, 5 mm]; 2) translation along direction $\mathbf{v}_G$ distributed uniformly on [−3 mm, 3 mm]; 3) rotation of the viewing direction $\mathbf{v}_G$ whose angle was distributed on $[0^\circ, 15^\circ]$ and rotation axis distributed uniformly on the $\mathbf{v}_G$ plane; 4) random roll applied about the perturbed direction vector $\mathbf{v}_G$, with angle distributed uniformly on $[-45^\circ, 45^\circ]$. This resulted in starting pose errors randomly distributed as follows: e_p uniformly distributed over [0, 5.8 mm], e_d uniformly distributed over $[0^\circ, 45^\circ]$, and e_r uniformly distributed over $[0^\circ, 45^\circ]$.

Table IV summarizes the results. The inverse-compositional method in general equalled or exceeded the accuracy of the forward-compositional method, with both IC and FC performing substantially better than NMI-based registration. In particular, the median errors for IC and FC were well under success criteria (29). Regarding registration success, IC proved to be the most robust method. In accordance with our discussion of Tables II and III, we emphasize that the registration success rates reported in Table IV tend to be conservative, as many registration successes arose from large perturbations.

Finally, using the patient-A data, we performed a second virtual-to-real test involving random perturbations of Θ_t over a large number of trials. We derived starting poses Θ_t by considering 450 random perturbations around Θ_G of each video frame. We computed the perturbations as above, but considered much wider starting error ranges: radial translation distributed uniformly on [0 mm, 10 mm], axial translation uniform on [−15 mm, 15 mm], directional rotation uniform on $[0^\circ, 20^\circ]$, and γ (roll) rotation uniform on $[-90^\circ, 90^\circ]$. From this candidate pose set, we selected starting poses Θ_t whose errors $\{e_p, e_d, e_r\}$ with respect to Θ_G fell within binned ranges. By independently varying each error metric separately, we computed success rates and registration errors as a function of initialization error.

Fig. 5 gives results. It is important to note that the perturbations of other pose parameters can be quite large for the results tabulated in Fig. 5. For example, a given trial for the $e_p = 2$ mm entry in Fig. 5 a can have a large additional rotation/roll error in accordance with the initial pose set described above. Nevertheless, IC and FC again both substantially outperformed the NMI-based method. The results also point out that initial position

difference is more important than initial direction difference for achieving successful registrations.

C. Continuous Registration and Computation Time

We next tested the ability of the proposed IC methodology (Algorithms 2–3) to perform continuous registration over a prerecorded video sequence. Hence, these tests simulated the ability of the proposed methods to track bronchoscope movement as the bronchoscope navigates through the airways. We again performed virtual-to-virtual and virtual-to-real tests.

For the virtual-to-virtual tests, we generated synthetic “video” sequences along two centerline-based airway paths in patient A, where a sequence consisted of MDCT-based VB views (Table V). One path started midway along the left main bronchus and proceeded eight airway generations through the left lung toward the left lower lobe. The second path started midway along the right main bronchus and traveled four airway generations through the right lung toward the right lower lobe. These idealized virtual-to-virtual tests again eliminated MDCT/video model differences. Also, because the paths followed smooth centerlines through the airways, they reduced the impact of non-ideal bronchoscope movements that occur during “real” bronchoscopy. Thus, they represented a best-case scenario for continuous registration.

Each test began with a starting pose Θ_i situated near the location of the first video frame. We then performed continuous registration over an entire sequence using Algorithms 2–3. For each video frame $\mathbf{I}_V(k)$, we compared our method’s optimized pose $\Theta_o(k)$ to the actual ground-truth pose of $\mathbf{I}_V(k)$, as given by the known pose of the VB view representing $\mathbf{I}_V(k)$. From these data, we computed the position error e_p and direction error e_d over an entire sequence.

Table V and Fig. 6 summarize the results. The proposed method successfully performed continuous registration over the entire sequences, with small mean registration errors. As the right-lung path was “straighter” than the left-lung path, the errors for the right lung tended to be smaller. As an indication of the quality of the registrations, the median position errors were under 0.5 mm, similar to the resolution of the underlying 3D MDCT scan. Errors arose from the pose estimate briefly lagging behind the true location within each airway, with large errors generally occurring near the transitional bifurcation regions.

It is important to note that registration succeeded over a sequence even if it “failed” at individual frames relative to the conservative criteria of (29). In fact, two related observations bear mentioning. First, registration succeeded for individual frames starting at locations having relatively large initial pose errors, with worst-case initial-pose errors being $e_p = 8$ mm and $e_d = 27^\circ$. Second, registration managed to “get back on track” after intermittent discrete registration “failures,” asserting the resilience and robustness of the registration process.

For the virtual-to-real tests, we drew upon six previously recorded bronchoscopic video sequences from the live human procedures associated with the Table I test cases. As these sequences arose from “real” bronchoscopy, the bronchoscope’s velocity varied between 0 and 20 mm/s during a sequence. Also, the sequences for patients B–D contained many more frames than those of patient A, because the bronchoscopist hovered longer over certain sites. After initializing a test at a sequence’s first video frame, we processed the sequence until an irrevocable failure occurred. We define an irrevocable failure as a situation whereby registration permanently loses track; i.e., it is a multi-frame event whereby the registered VB views clearly differed visually from the video frames over several consecutive frames. As a

rule, this situation implied that the registration process had irrevocably gone off track and could not recover during any subsequent frames [e.g., Fig. 7(d)].

Table VI summarizes the numerical results, Fig. 7 depicts sample registered image pairs and a registration failure, and Figs. 8 and 9 depict pictorial montages spanning complete sequences. Analogous the observations made earlier on how the registration process is able to get back on track after intermittent registration failures, Fig. 9 gives an example of a recovery after a registration failure.

These tests also point out that the registration success rates reported in Table IV in accordance with criteria (29) are overly conservative—registration success actually tends to be near 100% over the course of a continuous sequence. More importantly, the continuous process tends to readily “recover” from single-frame “misregistrations.” Along this line, it is vital to realize that single-frame misregistrations are fleeting transient events. During the course of a real-time sequence, such frames manifest themselves as mild “wobbles,” which become straightened out by subsequent frames. Such a phenomenon is not surprising, since the rapid 30 frames/s video rate implies that consecutive video frames, $\mathbf{I}_V(k)$ and $\mathbf{I}_V(k+1)$, only exhibit incremental changes, and the new incoming frames give opportunities for the registration process to get back on track. Nevertheless, registration failures can occur for small pose differences, as a small pose change can result in a large view change (as an analogy, envision suddenly approaching the edge of a cliff).

We also tested the robustness of NMI-based method for continuous registration for the sequences of patients C and D [8], [37]. The NMI method succeeded for the entire patient C sequence, but failed irrevocably by frame 29 for patient D. Furthermore, as discussed below, NMI-based registration unfortunately demands on the order of 8 s/frame, far below a real-time rate.

We point out that the IC methodology ran in real time for all sequences. Thus, it is clear that continuous real-time registration is feasible with the proposed methods. It is important, however, to measure the actual computation time required by the methods. The faster a method is, the better, as this would then free a guidance system to perform other tasks during a procedure.

To benchmark computational efficiency, we used a Dell Precision T7500 PC (2.8 GHz Xeon X5660 dual CPU, NVidia Quadro 4000 video card). We developed all computer code for the system using Visual C++, with little effort made to optimize code efficiency. For typical bronchoscopic video sequences, the IC method continuously processed video frames at a rate of 300 frames/s (3.3 ms/frame), with single-frame convergence generally occurring within 30 iterations—this rate is far beyond the real-time 30 frames/s video rate. As a comparison, tests with the forward method resulted in a rate of 0.096 frames/s (10.4 s/frame) [or 0.43 frames/s (2.33 s/frame) for $\varepsilon = 5 \times 10^{-4}$], while NMI could process 0.12 frames/s (8.51 s/frame) [8], [37]. Furthermore, the recent efforts by Deguchi and Luo quoted single-frame registration speeds on the order of 1 s/frame [39], [40]. In addition, the method of Rai, which interleaved OF-based video tracking and NMI-based single-frame registration, also achieved a rate on the order of 1 s/frame [41].

Hence, computational efficiency is a major advantage of the inverse-compositional formulation. The method not only allows true real-time processing during bronchoscopy, but also reduces the computational burden over other methods by two orders of magnitude.

D. Phantom Studies

The phantom studies considered the efficacy of the IC method over a series of live controlled bronchoscopy tests. These tests give an indication of the method's functionality during live real-time navigation, without the complications that occur during a live patient procedure such as breathing. Thus, they provide an indication of the method's potential for live interactive image-guided bronchoscopy.

For these studies, we used the MDCT data from patients A and E. In particular, for a given MDCT scan, we applied a series of previously validated automatic image-processing methods to segment the airway tree, extract the airway endoluminal surfaces, and define a series of navigation paths traversing the major airways [45], [46], [49]. Using the endoluminal surface data, we then fabricated two airway-tree phantoms out of acrylonitrile butadiene styrene, a common rigid thermoplastic used in rapid prototyping [9]. Thus, during live bronchoscopy, an airway-tree phantom served as “the patient.”

For all tests, we employed an Olympus BF P180 true-color bronchoscope (BF P180 scope #1 in Table I). Table VII specifies the paths navigated within the two phantoms. For a particular path, we first situated the bronchoscope at a proximal airway to initialize the system. We then navigated the bronchoscope along the path without stopping until an irrevocable registration failure occurred. The navigation times for the tests ranged from 7 to 14 s (30 frame/s video rate), the average bronchoscope velocity ranged from 8 to 14 mm/s, and the number of airways navigated ranged between three and eight generations. Figs. 10 and 11 give illustrative montages of registered views along two sample paths. The figures depict examples of the variety of maneuvers normally occurring during bronchoscopy, such as twisting and backing up. It is important to note, however, that since the phantoms consist of rigid airways involving no breathing motion, bronchoscopic navigation through a phantom is considerably easier than inside a real patient. Nevertheless, the tests clearly revealed the efficacy of the proposed IC method for real-time bronchoscopic navigation.

Note that path 1 of phantom A, navigated in 13 s, represented a remarkable trajectory spanning 8 airway generations. Navigation was relatively easy for this path, since it involved a “straight shot” from a generation-3 right middle-lobe airway to a peripheral generation-10 right lower-lobe airway trachea generation. In real patients, however, a physician essentially never navigates more than three consecutive airways before pausing. In addition, the physician as a rule carefully weighs maneuvers before making them and often hovers over certain sites before moving the bronchoscope—such motion patterns appeared prominently in the human video sequences considered earlier in Table VI.

Thus, neither the length of a video sequence nor the navigation time indicates how far the bronchoscope has moved. Using this observation, we believe that the number of airways traversed and the path length are the more telling limits on navigation success. By these measures, it is clear from the results of the phantom study and the earlier Section III-C tests that the IC method can operate in real-time continuously over substantial distances within the lungs.

E. System Integration and Human Studies

We have integrated the proposed registration methods into a complete planning and guidance system. The system facilitates both procedure planning and follow-on image-guided bronchoscopy for lung cancer staging. Here, we give a brief overview of the system and highlight its uses during bronchoscopy. Companion works give more complete description [45], [46], [49]–[51].

During procedure planning, the system provides tools for extracting the airway tree, defining diagnostic targets, and computing suitable centerline paths leading to the defined targets. This results in a procedure plan employed during image-guided bronchoscopy. To apply the system during bronchoscopy for a particular target, the system first cues the physician with a VB view near the main carina. The physician then navigates the bronchoscope close to this view. At this point, system guidance begins in two modes: 1) single-view registration at each major bifurcation along a guidance path; or 2) continuous registration. For the single-view mode, registration occurs instantaneously, thanks to the rapid convergence of the inverse-compositional method. By comparison, previous work drawing upon NMI-based single-frame registration required several seconds per registration [8], [37]. (To the best of our knowledge, these are the only other prior works that have applied automatic computer-based CT-video registration to human patients during live bronchoscopy.)

Table VIII profiles a 10-patient pilot study focusing on mediastinal nodal staging. The study was conducted under IRB approval at our University's medical center, with all patients providing informed consent. The first two cases established system proof of concept and helped determine the ergonomic factors in applying the system clinically. Example factors included understanding proper work flow from beginning to end, finalizing system configuration in the bronchoscopy suite, and establishing smooth communication with the attending staff during a procedure. (Patient 1 corresponds to patient A in Table I.) For the remaining patients, we entered a "production mode," whereby all planning and guidance procedures were run in the context of the lung cancer clinic.

Fig. 12 gives example system views for patient 8, who underwent nodal staging. The figure illustrates successful guidance and final biopsy of a hilar node in station 10L. During bronchoscopy, the system keeps all views continuously synchronized. It also employs data fusion, whereby information from the MDCT-based procedure plan is graphically merged with the video. Fig. 13 gives sample views for patient 6, which involved more difficult peripheral maneuvers. For patient 7, the physician decided to terminate the procedure, as the patient's oxygen level dropped to an unacceptable level while under sedation. Guidance did begin successfully, however, for this case. Overall, the study showed the system's feasibility in the clinical environment. The number of airways navigated for the 19 targets considered ranged from 2 to 5 generations. In all cases, the system functioned smoothly and safely and provided useful guidance to the predesignated diagnostic targets.

Our system has also been successfully evaluated in a phantom study for peripheral nodule biopsy. The study, which involved 12 physicians, compared the guidance system to standard bronchoscopy practice. Results showed that image-based guidance significantly improves peripheral-lung biopsy performance over standard practice from 43% to 94% [9]. In addition, several other studies have shown that the system fits smoothly within the standard two-step lung cancer staging work flow [46], [50], [51], [68]. For example, we have applied the system to 30 patients who underwent peripheral bronchoscopy (age range: 18–85 years; 9 males, 21 females). For a typical peripheral target, the physician navigated an ultrathin bronchoscope 7.5 airways and performed the first target biopsy in 6 min 48 s (time includes system initialization and localization of biopsy site) [51], [68]. This compares very favorably to a study using a competing system (mean route length = 5.7 airways, mean time to first sample = 8 min 30s [69]). During this study, we noted that for one patient, cough and extreme respiratory variation made sampling difficult, while for another patient, laryngospasm forced early procedure termination. Yet, for all patients, the system functioned properly with no adverse system events.

IV. Discussion

The inverse-compositional framework provides a robust accurate method for CT-video registration in image-guided bronchoscopy. In addition, instead of taking seconds per frame, the method accomplishes single-frame registration in milliseconds. Hence, as demonstrated for both video sequences and live clinical situations, truly interactive image-guided bronchoscopy can occur at real-time frame rates. On another front, the registration method, which draws upon the concept of image-based rendering for video warping, is general, and, hence, could conceivably be applied to other image-guided endoscopy applications such as colonoscopy [70], [71].

An alternative forward-compositional approach uses no precomputation step, requires no storage of precomputed data, and avoids video warping. We do not recommend the method, however, because it offers no advantage in robustness or accuracy and demands two orders of magnitude more computation than the inverse approach. Also, as has been reported previously in the field of image alignment, the forward and inverse compositional methods, while approximately equivalent, do not strictly speaking converge to the same results [56]. It is important to consider the forward method, however, given its potential advantages. As future work, we note that the inverse-compositional method could conceivably be improved by employing enhancements offered by the Levenberg–Marquardt algorithm.

The inverse method does require high-quality VB views to be effective (as does the forward method). This entails having effective MDCT-based airway-tree segmentation and endoluminal surface generation. Over a wide range of human tests, our methods for these operations have shown that successful registration can occur at locations deeper than generation-10 airways [45], [46], [50]. Conversely, we point out that registration failures can occur early in the airway tree, but the overarching factor in such failures is the starting pose, not the quality of the VB views. In addition, registration performance depends on MDCT spatial resolution. For our studies where voxel resolution was on the order of 0.5–1.0 mm, registration is largely successful for airways having physical diameter > 2 mm ($\approx$ generation 7 or less). The generation-11 site of Fig. 10(g) illustrates this situation, where the registration failure can be attributed to the small-airway size—note the blocky-looking VB view, a tell-tail sign of insufficient resolution.

The method has been integrated successfully into a system for image-guided bronchoscopy. Human studies indicate that the system can fit seamlessly into the clinical work flow and outperforms competing systems. Deguchi and Luo, however, pointed out the prevalence of registration failures for video frames corrupted by air bubbles and specularities and for frames exhibiting low-information content [39], [40]. Note that our methods often succeed for such frames; e.g., Fig. 4(b) and Fig. 13(b). As stated in Section II-C, our method achieves a mild level of immunity from such imperfections, because the registration calculation focuses on the region about airway bifurcations. Fig. 9(c), however, shows a failure that occurred at a frame excessively corrupted by air bubbles. Our methods can also fail for low-information frames, such as those where the bronchoscope points at an airway wall. But, even here, if the event is “short” (i.e., involves few video frames), then the method can recover and get back on track, similar to the results discussed in Section III-C.

Occlusions also could possibly result in registration errors. We believe, however, that occlusions, if any, have little impact given that the airways have a largely convex structure and that the actual distances considered are small. Furthermore, the bronchoscope in general moves slowly through the airway structure relative to the video frame rate. Hence, the video source is highly oversampled in time (i.e., highly redundant), implying that the net positional change—and overall change in the view—from one frame to the next tends to be small. Our

results for video sequences and for phantom and human studies appear to bear these points out.

The biggest difficulty arises when a patient coughs during a procedure. Such episodes can last for many seconds. To account for such procedural events, our complete guidance system employs a fail-safe guidance feature referred to as the “base camp,” whereby successive bronchoscope maneuvers are captured and saved (generally the previously encountered bifurcation) [50]. With this strategy, after a coughing episode, the physician can back up to last known correct location (the “base camp”) and reinitialize guidance at this location.

Acknowledgments

The authors would like to thank P. Byrnes for his help with the preparation of the manuscript and the generated results.

This work was supported by the National Institutes of Health (NIH) National Cancer Institute under Grant R01-CA074325, Grant R01-CA151433, and Grant R44-CA091534. W. E. Higgins has an identified conflict of interest related to Grant R01-CA151433, which is under management by Penn State and has been reported to the NIH.

References

1. Siegel R, Naishadham D, Jemal A. Cancer statistics 2012. *CA Cancer J Clin.* Jan; 2012 62(1):10–29. [PubMed: 22237781]
2. Sihoe AD, Yim AP. Lung cancer staging. *J Surg Res.* Mar; 2004 117(1):92–106. [PubMed: 15013719]
3. Wang, KP.; Mehta, AC. Flexible Bronchoscopy. 2. Turner, JF., editor. Cambridge, MA: Blackwell; 2003.
4. Ueno J, Murase T, Yoneda K, Tsujikawa T, Sakiyama S, Kondoh K. Three-dimensional imaging of thoracic diseases with multi-detector row CT. *J Med Invest.* Aug; 2004 51(3–4):163–170. [PubMed: 15460902]
5. Dolina MY, Cornish DC, Merritt SA, Rai L, Mahraj R, Higgins WE, Bascom R. Interbronchoscopist variability in endobronchial path selection: A simulation study. *Chest.* Apr; 2008 133(4):897–905. [PubMed: 18263679]
6. Asano F. Virtual bronchoscopic navigation. *Clin Chest Med.* Mar; 2010 31(1):75–85. [PubMed: 20172434]
7. Minami H, Ando Y, Nomura F, Sakai S, Shimokata K. Inter-bronchoscopist variability in the diagnosis of lung cancer by flexible bronchoscopy. *Chest.* Jun; 1994 105(2):1658–1662. [PubMed: 8205857]
8. Helferty JP, Sherbondy AJ, Kiraly AP, Higgins WE. Computer-based system for the virtual-endoscopic guidance of bronchoscopy. *Comput Vis Image Understand.* Oct-Nov; 2007 108(1–2): 171–187.
9. Merritt SA, Gibbs JD, Yu KC, Patel V, Rai L, Cornish DC, Bascom R, Higgins WE. Real-time image-guided bronchoscopy for peripheral lung lesions: A phantom study. *Chest.* Nov; 2008 134(5):1017–1026. [PubMed: 18583513]
10. Tsushima K, Sone S, Hanaoka T, Takayama F, Honda T, Kubo K. Comparison of bronchoscopic diagnosis for peripheral pulmonary nodule under fluoroscopic guidance with CT guidance. *Respir Med.* Apr; 2006 100(4):737–745. [PubMed: 16125921]
11. Yasufuku K, et al. Endobronchial ultrasound guided transbronchial needle aspiration for staging of lung cancer. *Lung Cancer.* Dec; 2005 50(3):347–354. [PubMed: 16171897]
12. Geraghty PR, Kee ST, McFarlane G, Razavi MK, Sze DY, Dake MD. CT-guided transthoracic needle aspiration biopsy of pulmonary nodules: Needle size and pneumothorax rate. *Radiology.* Nov; 2003 229(2):475–481. [PubMed: 14595149]
13. Vining DJ, Liu K, Choplin RH, Haponik EF. Virtual bronchoscopy: Relationships of virtual reality endobronchial simulations to actual bronchoscopic findings. *Chest.* Feb; 1996 109(2):549–553. [PubMed: 8620734]

14. Summers RM. Navigational aids for real-time virtual bronchoscopy. *Am J Roentgen*. May; 1997 168(5):1165–1170.
15. Rapp-Bernhardt U, Welte T, Budinger M, Bernhardt TM. Comparison of three-dimensional virtual endoscopy with bronchoscopy in patients with oesophageal carcinoma infiltrating the tracheobronchial tree. *Br J Radiol*. Dec.1998 71:1271–1278. [PubMed: 10319000]
16. Higgins WE, Ramaswamy K, Swift R, McLennan G, Hoffman EA. Virtual bronchoscopy for 3D pulmonary image assessment: State of the art and future needs. *Radiographics*. May-Jun;1998 18(3):761–778. [PubMed: 9599397]
17. McAdams HP, Goodman PC, Kussin P. Virtual bronchoscopy for directing transbronchial needle aspiration of hilar and mediastinal lymph nodes: A pilot study. *Am J Roentgenol*. May; 1998 170(5):1361–1364. [PubMed: 9574616]
18. Haponik EF, Aquino SL, Vining DJ. Virtual bronchoscopy. *Clin Chest Med*. Mar; 1999 20(1):201–217. [PubMed: 10205726]
19. Hopper K, Lucas T, Gleeson K, Stauffer J, Bascom R, Mauger D, Mahraj R. Transbronchial biopsy with virtual CT bronchoscopy and nodal highlighting. *Radiology*. Nov; 2001 221(2):531–536. [PubMed: 11687700]
20. Shinagawa N, Yamazaki K, Onodera Y, Miyasaka K, Kikuchi E, Dosaka-Akita H, Nishimura M. CT-guided transbronchial biopsy using an ultrathin bronchoscope with virtual bronchoscopic navigation. *Chest*. Mar; 2004 125(3):1138–1143. [PubMed: 15006979]
21. Weiner GM, Schulze K, Geiger B, Ebhardt H, Wolf K, Albrecht T. CT bronchoscopic simulation for guiding transbronchial needle aspiration of extramural mediastinal and hilar lesions: Initial clinical results. *Radiology*. Mar; 2009 250(3):923–931. [PubMed: 19244055]
22. Rubin GD, Beaulieu CF, Argiro V, Ringl H, Norbash AM, Feller JF, Dake MD, Napel S, Jeffrey RB, Napel S. Perspective volume rendering of CT and MR images: Applications for endoscopic imaging. *Radiology*. May; 1996 199(2):321–330. [PubMed: 8668772]
23. Rogalla, P.; Van Scheltinga, J.; Hamm, B. *Virtual Endoscopy and Related 3D Techniques*. Berlin, Germany: Springer-Verlag; 2002.
24. Deligianni F, Chung AJ, Yang G. Nonrigid 2-D/3-D registration for patient specific bronchoscopy simulation with statistical shape modeling: Phantom validation. *IEEE Trans Med Imag*. Nov; 2006 25(11):1462–1471.
25. Solomon SB, White P Jr, Wiener CM, Orens JB, Wang KP. Three-dimensional CT-guided bronchoscopy with a real-time electromagnetic position sensor: A comparison of two image registration methods. *Chest*. Dec; 2000 118(6):1783–1787. [PubMed: 11115473]
26. Gildea TR, Mazzone PJ, Karnak D, Meziane M, Mehta AC. Electromagnetic navigation diagnostic bronchoscopy: A prospective study. *Am J Resp Crit Care Med*. Nov; 2006 174(9):982–989. [PubMed: 16873767]
27. Schwarz Y, Greif J, Becker HD, Ernst A, Mehta A. Real-time electromagnetic navigation bronchoscopy to peripheral lung lesions using overlaid CT images: The first human study. *Chest*. Apr; 2006 129(4):988–994. [PubMed: 16608948]
28. Eberhardt R, Anantham D, Ernst A, Feller-Kopman D, Herth F. Multimodality bronchoscopic diagnosis of peripheral lung lesions: A randomized controlled trial. *Am J Respir Crit Care Med*. Jul; 2007 176(1):36–41. [PubMed: 17379850]
29. Gergel, I.; dos Santos, TR.; Tetzlaff, R.; Maier-Hein, L.; Meinzer, H-P.; Wegner, I. Particle filtering for respiratory motion compensation during navigated bronchoscopy. In: Wong, KH.; Miga, MI., editors. *SPIE Medical Imaging 2010: Visualization, Image-Guided Procedures, and Modeling*. Vol. 7625. 2010. p. 76250W1-76250W8.
30. Luo, X.; Feuerstein, M.; Sugiura, T.; Kitasaka, T.; Imaizumi, K.; Hasegawa, Y.; Mori, K. Towards hybrid bronchoscope tracking under respiratory motion—Evaluation on a dynamic motion phantom. In: Wong, KH.; Miga, MI., editors. *SPIE Medical Imaging 2010: Visualization, Image-Guided Procedures, and Modeling*. Vol. 7625. 2010. p. 76251B1-76251B12.
31. Soper TD, Haynor DR, Glenny RW, Seibel EJ. In vivo validation of a hybrid tracking system for navigation of an ultrathin bronchoscope within peripheral airways. *IEEE Trans Biomed Eng*. Mar; 2010 57(3):736–745. [PubMed: 19846362]

32. Grimson WE, Ettinger GJ, White SJ, Lozano-Perez T, Wells WE III, Kikinis R. An automatic registration method for frameless stereotaxy, image guided surgery, and enhanced reality visualization. *IEEE Trans Med Imag.* Apr; 1996 15(2):129–140.
33. Maurer CR Jr, Fitzpatrick JM, Wang MY, Galloway RL Jr, Maciunas RJ, Allen GS. Registration of head volume images using implantable fiducial markers. *IEEE Trans Med Imag.* Aug; 1997 16(4):447–462.
34. Stefansic JD, Herline AJ, Shyr Y, Chapman WC, Fitzpatrick JM, Dawant BM, Galloway RL Jr. Registration of physical space to laparoscopic image space for use in minimally invasive hepatic surgery. *IEEE Trans Med Imag.* Oct; 2000 19(10):1012–1023.
35. Bricault I, Ferretti G, Cinquin P. Registration of real and CT-derived virtual bronchoscopic images to assist transbronchial biopsy. *IEEE Trans Med Imag.* Oct; 1998 17(5):703–714.
36. Helferty JP, Zhang C, McLennan G, Higgins WE. Videoendoscopic distortion correction and its application to virtual guidance of endoscopy. *IEEE Trans Med Imag.* Jul; 2001 20(7):605–617.
37. Higgins WE, Helferty JP, Lu K, Merritt SA, Rai L, Yu KC. 3D CT-video fusion for image-guided bronchoscopy. *Comput Med Imag Graph.* Apr; 2008 32(3):159–173.
38. Mori K, Deguchi D, Sugiyama J, Suenaga Y, Toriwaki J, Maurer CR, Takabatake H, Natori H. Tracking of bronchoscope using epipolar geometry analysis and intensity-based image registration of real and virtual endoscopic images. *Med Image Anal.* Sep; 2002 6(3):321–336. [PubMed: 12270236]
39. Deguchi D, Mori K, Feuerstein M, Kitasaka T, Maurer CR, Suenaga Y, Takabatake H, Mori M, Natori H. Selective image similarity measure for bronchoscope tracking based on image registration. *Med Image Anal.* Jul; 2009 13(1):621–633. [PubMed: 19592291]
40. Luo, X.; Feuerstein, M.; Reichl, T.; Kitasaka, T.; Mori, K.; Edwards, PJ.; Liao, H.; Pan, X., editors. 5th Int Workshop Med Imag Augmented Reality. 2010. An application driven comparison of several feature extraction algorithms in bronchoscope tracking during navigated bronchoscopy; p. 475–484.vol. LNCS 6326
41. Rai L, Helferty JP, Higgins WE. Combined video tracking and image-video registration for continuous bronchoscopic guidance. *Int J Comput Assist Radiol Surg.* Sep; 2008 3(3–4):315–329.
42. Rai L, Merritt SA, Higgins WE. Real-time image-based guidance method for lung-cancer assessment. *IEEE Conf Comput Vis Pattern Recognit.* 2006:2437–2444.
43. Lowe DG. Distinctive image features from scale-invariant key-points. *Int J Comput Vis.* 2004; 60(2):91–110.
44. Szeliski, R. *Computer Vision—Algorithms and Applications*. New York, NY: Springer; 2011.
45. Gibbs JD, Graham MW, Higgins WE. 3D MDCT-based system for planning peripheral bronchoscopic procedures. *Comput Biol Med.* Mar; 2009 39(3):266–279. [PubMed: 19217089]
46. Graham MW, Gibbs JD, Cornish DC, Higgins WE. Robust 3D airway-tree segmentation for image-guided Peripheral Bronchoscopy. *IEEE Trans Med Imag.* Apr; 2010 29(4):982–997.
47. Kiraly AP, Hoffman EA, McLennan G, Higgins WE, Reinhardt JM. 3D human airway segmentation methods for clinical virtual bronchoscopy. *Acad Radiol.* Oct; 2002 9(10):1153–1168. [PubMed: 12385510]
48. Kiraly AP, Helferty JP, Hoffman EA, McLennan G, Higgins WE. 3D path planning for virtual bronchoscopy. *IEEE Trans Med Imag.* Nov; 2004 23(11):1365–1379.
49. Yu KC, Gibbs JD, Graham MW, Higgins WE. Image-based reporting for bronchoscopy. *J Digital Imag.* Feb; 2010 23(1):39–50.
50. Graham MW, Gibbs JD, Higgins WE. Computer-based route-definition system for peripheral bronchoscopy. *J Digital Imag.* Apr; 2012 25(2):307–317.
51. Gibbs JD, Graham MW, Bascom R, Cornish DC, Khare R, Higgins WE. Computer-based planning and guidance system for image-guided bronchoscopy. *IEEE Trans Biomed Eng.* submitted for publication.
52. Lucas B, Kanade T. An iterative image registration technique with an application to stereo vision. *Proc Int Joint Conf Artif Intell.* 1981:674–679.
53. Hager GD, Belhumeur PN. Efficient region tracking with parametric models of geometry and illumination. *IEEE Trans Pattern Anal Mach Intell.* Oct; 1998 20(10):1025–1039.

54. Shum HY, Szeliski R. Construction of panoramic image mosaics with global and local alignment. *Int J Comput Vis.* Feb; 2000 16(1):63–84.
55. Baker S, Matthews I. Equivalence and efficiency of image alignment algorithms. *Proc IEEE Conf Comput Vis Pattern Recognit.* 2001:1090–1097.
56. Baker S, Matthews I. Lucas-Kanade 20 years on: A unifying framework. *Int J Comput Vis.* Feb; 2004 56(3):221–255.
57. Evangelidis GD, Psarakis EZ. Parametric image alignment using enhanced correlation coefficient maximization. *IEEE Trans Pattern Anal Mach Intell.* Oct; 2008 30(10):1858–1865. [PubMed: 18703836]
58. Izquierdo E, Ohm JR. Image-based rendering and 3D modeling: A complete framework. *Signal Process: Image Commun.* Aug; 2000 15(10):817–858.
59. Trucco, E.; Verri, A. *Introductory Techniques for 3-D Computer Vision.* Upper Saddle River, NJ: Prentice-Hall; 1998.
60. Ma, Y.; Kosecka, J.; Soatto, S.; Sastry, S. *An Invitation to 3-D Vision—From Images to Models.* New York: Springer; 2001.
61. Okatani T, Deguchi K. Shape reconstruction from an endoscope image by shape from shading technique for a point light source at the projection center. *Comput Vis Image Understand.* May; 1997 66(2):119–131.
62. Baker, S.; Matthews, I. Tech Rep CMU-RI-TR-04-64. 2004. Lucas-Kanade 20 years on: Part 5 Carnegie Mellon Univ. Technical Report CMU-RI-TR-04-64
63. Romdhani S, Vetter T. Efficient, robust and accurate fitting of a 3D morphable model. *Proc 9th IEEE Int Conf Comput Vis.* 2003; 1:59–66.
64. Baker S, Kanade T. Limits on super-resolution and how to break them. *IEEE Conf Comput Vis Pattern Recognit.* Jun 15.2000 2:372–379.
65. Flood, B.; Rai, L.; Higgins, WE. System for robust bronchoscopic video distortion correction. In: Wong, KH.; Holmes, DR., editors. *SPIE Medical Imaging 2011: Visualization, Image-Guided Procedures, and Modeling.* Vol. 7964. 2011. p. 79641L-1-79641L-16.
66. Hearn, D.; Baker, MP. *Computer Graphics with OpenGL.* 3. Upper Saddle River, NJ: Pearson Prentice Hall; 2004.
67. Merritt, SA.; Rai, L.; Higgins, WE. Real-time CT-video registration for continuous endoscopic guidance. In: Manduca, A.; Amini, AA., editors. *SPIE Medical Imaging 2006: Physiology, Function, and Structure from Medical Images.* 2006. p. 370-384.
68. Higgins WE, Bascom R, Graham MW, Gibbs JD, Cornish DC, Khan M, Khare R. Image-guided bronchoscopic sampling of peripheral lesions: A human study. *Am J Respirat Crit Care Med.* May. 2010 181:A5171–A5171.
69. Shinagawa N, et al. Virtual bronchoscopic navigation system shortens the examination time: Feasibility study of virtual bronchoscopic navigation system. *Lung Cancer.* May; 2007 56(2):201–206. [PubMed: 17229486]
70. Summers RM, Hara A, Luboldt W, Johnson C. Computed tomographic and magnetic resonance colonography: Sum summary of progress from 1995 to 2000. *Current Problems Diagnostic Radiol.* Sep-Oct;2001 30(5):141–168.
71. Liu J, Subramanian KR, Yoo TS. Region flow: A multi-stage method for colonoscopy tracking. *Proc 13th Int Conf Med Image Comput Comput-Assist Intervent: Part II.* 2010:505–513.

Appendix

We can also solve (15) by using a forward-compositional formulation. Following a derivation similar to the one given in Section II-B for the inverse-compositional formulation, we instead minimize the cost (30)

$$\hat{C}(\mathbf{I}_{CT}, \mathbf{I}_V, \Theta) = \sum_{u,v} \left[\bar{\mathbf{I}}_{CT}^{(\Theta \circ \Delta \Theta)}(u, v) - \bar{\mathbf{I}}_V(u, v) \right]^2 \quad (30)$$

with respect to $\Delta\Theta$, where $(\Theta \circ \Delta\Theta)$ signifies the composition of two mappings for $\bar{\mathbf{I}}_{\text{CT}}^{(\Theta \circ \Delta\Theta)}$ via (9) and (10). At each iteration, we compute the pose update

$$\mathbf{M}_{\Theta} \leftarrow \mathbf{M}_{\Theta} \cdot \mathbf{M}_{\Delta\Theta} \quad (31)$$

with

$$\Delta\Theta = \mathbf{H}^{-1} \cdot \sum_{u,v} [\nabla \bar{\mathbf{I}}_{\text{CT}}]^T \cdot [\bar{\mathbf{I}}_V(u,v) - \bar{\mathbf{I}}_{\text{CT}}^{\Theta}(u,v)] \quad (32)$$

where ∇_{CT} and $\mathbf{H}^{-1}$ are given by (23) and (25). Thus, similar to the inverse method, which iterates (21) and (24) until (20) converges, the forward method iterates (31) and (32) until (30) converges. Upon convergence, the optimal transformation enables the virtual and real spaces to become registered per (11), where $\bar{\mathbf{I}}_{\text{CT}}^{\Theta_o}$ approximately matches $\bar{\mathbf{I}}_V$.

Implementation of the forward method is similar to Algorithm 3, but involves no precomputation step as with Algorithm 1 and no storage of precomputed data. Unfortunately, at *every* iteration, the quantities $\{\bar{\mathbf{I}}_{\text{CT}}, \nabla_{\text{CT}}, \mathbf{H}^{-1}\}$ must be recomputed in (32) for the updated pose Θ in (31). This adds an especially heavy computational burden, since, from (23) and (27), the computation of ∇_{CT} alone demands 12 new VB-views (two views per gradient component).

Fig. 1.

Illustration of the virtual and real 3D spaces involved in the registration problem for image-guided bronchoscopy. The patient's chest corresponds to the World imaged by these two spaces. (a) 3D MDCT-based virtual space at a pose Θ , where the left view depicts the patient's 3D MDCT scan (defines the complete virtual space), the middle view shows the rendered airway tree segmented from the MDCT scan, and the right view depicts a local VB view I_{CT}^Θ at pose Θ in the virtual space; the airway tree shows the virtual bronchoscopic camera at pose Θ . (b) Real space, where the left view depicts a rendition of the bronchoscope's position (real camera) at a pose inside the patient's airway tree and the right view shows a corresponding bronchoscopic video frame I_V (left view adapted from [3, Fig. 19.4]). Both the VB view I_{CT}^Θ and video frame I_V shown in the figure are near the main carina, but are not registered. (a) MDCT-based virtual space. (b) Bronchoscopy-based real space.

Fig. 2.

Example VB view and associated gradient-component images ∇_{CT} per (23) and (27). Top row of (b) corresponds to the angular components $\{ \nabla_{CT}/\alpha, \nabla_{CT}/\beta, \nabla_{CT}/\gamma \}$, while the bottom row corresponds to the translational components $\{ \nabla_{CT}/t_x, \nabla_{CT}/t_y, \nabla_{CT}/t_z \}$. (a) VB view I_{CT} . (b) Gradient vector ∇_{CT} .

Fig. 3.

Example of image-based rendering during inverse-compositional registration. The figure depicts the operations occurring during the first iteration of lines 5–7 of Algorithm 3. (a) Input video frame $\mathbf{I}_V$. (b) VB view $\mathbf{I}_{CT}^{\Theta_t}$ at target global pose Θ_t . (c) $\mathbf{I}_{CT}^{\Theta_t}$ with its 2-D (u, v) grid superimposed; a 16×16 grid is shown for clarity—the actual grid has the same resolution as $\mathbf{I}_{CT}^{\Theta_t}$. (d) Relative warp $\mathbf{W}$ between Θ_t and best reference pose Θ_i applied to 2D grid of (c) per line 6 of Algorithm 3—the warped grid is superimposed on $\mathbf{I}_V$. (e) Warp applied to $\mathbf{I}_V$, resulting in $\mathbf{I}_V^{\Theta}(u, v)$, with grid superimposed—the warp relating the two grids in (c) and (d) is used to interpolate values for the warped video frame. (f) Warped video frame of (e), $\mathbf{I}_V^{\Theta}(u, v)$, without grid.

Fig. 4.

Example ground-truth data used in the single-frame registration tests. The two pairs are from patient A. In each pair, the left view is the target video frame $\mathbf{I}_R$, while the right view is the corresponding manually derived ground-truth VB view $\mathbf{I}_{CT}^{\Theta_G}$. (a) Distal end of trachea. (b) Distal left main bronchus toward left upper lobe.

Fig. 5.

Virtual-to-real single-frame registration success versus initial pose error. Video frames for patient A used. Each data point represents the success rates computed from those registration results whose initial pose error is as indicated. Registration success was defined by the stringent bounds of (29). For each plot, IC = solid line with $\blacklozenge$ data points, FC = “—..—.” line and $\circ$ points, NMI = “—..—.” line and $\times$ points. (a) Success rate versus initial position error e_p ; initial pose Θ_r also satisfied $e_d < 15^\circ$ and $e_r < 22.5^\circ$. (b) Success rate versus initial direction error e_d ; initial pose also satisfied $e_p < 7.5$ mm and $e_r < 22.5^\circ$. (c) Success rate versus initial roll error e_r ; initial pose also satisfied $e_p < 7.5$ mm and $e_d < 15^\circ$. (a) Success Rate versus initial position error e_p . (b) Success rate versus initial direction error e_d . (c) Success rate versus initial roll error e_r .

Fig. 6.

Virtual-to-virtual continuous registration tests [67]. The figure presents error metric plots for the airway paths listed in Table V. (a) Path through left lung. (b) Path through right lung. The three plots associated with each path are as follows. Left plot: distance traveled as a function of frame number for both the ground-truth position and registered position. Middle plot: position error versus frame number. Right plot: direction error versus frame number. Note that the distance traveled along a path is not strictly linear relative to frame number. (a) Left lung path. (b) Right lung path.

Fig. 7.

Sample registered virtual/real image pairs ($\mathbf{I}_{CT}^{\Theta_o}(k)$, $\mathbf{I}_V(k)$) for the continuous virtual-to-real registration tests [67]. (a) Patient A, sequence 1, frame $k = 22$. (b) Patient B, sequence 2, $k = 11$. (c) Patient B, sequence 2, $k = 395$. (d) Registration failure: patient B, sequence 2, $k = 615$ —registrations for all subsequent frames reveal that the process has gone off track.

Fig. 8.

Sample registered virtual/real image pairs ($\mathbf{I}_V(k), \mathbf{I}_{CT}^{\Theta_o}(k)$) spanning the patient-C continuous registration test. (a) Frame $k = 1$. (b) $k = 89$. (c) $k = 405$. (d) $k = 858$. (e) $k = 1,049$. (f) $k = 1,907$.

Fig. 9.

Sample registered virtual/real image pairs ($\mathbf{I}_V(k)$, $\mathbf{I}_{CT}^{\Theta}(k)$) spanning the patient-D continuous registration test. Registration “fails” for frame $k = 796$ (c), but successfully recovers to get back on track [e.g., (d)]. (a) Frame $k = 1$. (b) $k = 131$. (c) $k = 796$ (failure). (d) $k = 875$.

Fig. 10.

Sample registered virtual/real image pairs ($\mathbf{I}_V(k)$, $\mathbf{I}_{CT}^{\Theta_o}(k)$) spanning path 1 of phantom A. For each view, k equals the video frame number, while “gen.” equals the airway generation number along the path. Part (g) gives an example frame *after* the irrevocable failure (on frame $k = 393$) for this path. (a) $k = 1$, gen. 3. (b) $k = 68$, gen. 4. (c) $k = 124$, gen. 5. (d) $k = 196$, gen. 7. (e) $k = 319$, gen. 8. (f) $k = 363$, gen. 9. (g) $k = 397$, gen. 11. (fail).

Fig. 11.

Sample registered virtual/real image pairs ($\mathbf{I}_V(k)$, $\mathbf{I}_{CT}^{\Theta_o}(k)$) spanning path 3 of phantom E. This path begins at a proximal location in right main bronchus (a) and continues to a generation-5 right-lung airway. Note that the user backed up between sites (e) and (f). Thus, views (d) and (f) depict the proximal and distal ends, respectively, of the generation-4 airway traversed by this path. (a) $k = 1$, gen. 2. (b) $k = 62$, gen. 3. (c) $k = 113$, gen. 3. (d) $k = 209$, gen. 4. (e) $k = 288$, gen. 5. (f) $k = 384$, gen. 4.

Fig. 12.

Example system views for a 44-year-old female during a lymph-node staging procedure (patient 8). Part (a) depicts the system's display at a particular site en route 22.3 mm from the center of a station 10L lymph node. The top section shows a registered pair for the current video frame (left) and VB view (right). The bottom section depicts the rendered airway tree (left) and axial 2D MDCT section $z = 205$ (right) corresponding to the current pose, with the cross-hairs indicating the specific $x - y$ location. On all views, the red region represents the target node, red lines indicate airway centerlines, and the blue line represents the guidance path leading to the node. By using data fusion, the guidance path and predefined lymph node are superimposed on the video. Regarding the rendered airway tree, the green needle on the blue guidance path indicates the bronchoscope's current pose; note how the needle points toward the node. Part (b) depicts a registered pair at the time of biopsy 17.5 mm from the node's center—note the appearance of the biopsy needle in the video. (MDCT scan characteristics: voxel resolution, $\Delta x = \Delta y = 0.65$ mm, $\Delta z = 0.5$ mm; slice thickness = 0.75 mm; scan dimensions, $656\ 512 \times 512$ 2D axial-plane sections). (a) System display during image-guided bronchoscopy. (b) Registered pair during biopsy (needle visible in video).

Fig. 13.

Example system views at two synchronized sites for a 50-year-old male during a lymph-node staging procedure (patient 4). (MDCT scan characteristics: voxel resolution, $\Delta x = \Delta y = 0.82$ mm, $\Delta z = 0.5$ mm; slice thickness = 0.75 mm; scan dimensions, $675 \times 512 \times 512$ 2D axial-plane sections.) For each system display, we depict a weighted-sum coronal projection of the patient's 3D MDCT chest scan (left view) along with a registered video/VB-view pair (right view). The projection views, which resemble standard chest X-rays, depict the guidance route (blue line) and progress made thus far along the route (yellow line), while the red lines denote the other airway centerlines. Note the successful registration for the video frame of (b) despite the bubbles. We again fuse the guidance paths and diagnostic targets onto the video views. (a) Display at generation-4 for right-upper lobe target region (green). (b) Display at generation-5 for a right-middle lobe target (blue).

Test Cases Used in Sections III-B–III-D. All 3D MDCT Scans Consisted of $Z\ 512 \times 512$ 2D Sections, With Each Section Having Voxel Resolution ($\Delta x, \Delta y, \Delta z$). Bronchoscopes Produced Video at a Rate of 30 frames/s

TABLE I

Patient	MDCT Scan Specifications			Bronchoscopic Video Specifications	
	No. of Sections Z	($\Delta x, \Delta y, \Delta z$) (mm)	Scanner	Video-Frame Size	Bronchoscope
A	706	0.67, 0.67, 0.5	Siemens Sensation-16	592×440 (4)	Pentax EB 1570-K
B	574	0.72, 0.72, 0.6	Marconi Mx8000	264×264 (2)	Olympus BF200
C	590	0.70, 0.70, 0.5	Siemens Sensation-40	500×470 (4)	Olympus BF P180, #1
D	562	0.70, 0.70, 0.5	Siemens Sensation-40	544×605 (4)	Olympus BF P180, #2
E	373	0.80, 0.80, 0.7	Philips Gemini True Flight	500×470 (4)	Olympus BF P180, #1

Parameter Sensitivity Results for Single-Frame Registration. Test Data for Patients A, B, and C Used. For Each Parameter of Starting Pose Θ_I , the Ranges Indicate the Minimum and Maximum Perturbations From Θ_G That Enabled Successful Registration Over All Test Frames Per Criteria (29) (a) Virtual-to-Virtual Results (b) Virtual-to-Real Results

TABLE II

(a)						
Method	Position ranges (mm)			Angular ranges (degrees)		
	t_x	t_y	t_z	α	β	γ
IC	-2.8, 3.2	-3.8, 3.6	-6.4, 6.2	-23.0, 32.0	-23.2, 30.0	-50.4, 43.6
FC	-2.8, 3.4	-3.8, 3.6	-5.5, 5.6	-16.4, 16.8	-14.0, 14.8	-19.6, 20.8
NMI	-2.4, 2.6	-3.3, 3.2	-6.0, 4.4	-16.4, 7.0	-14.0, 8.0	-22.5, 14.0
(b)						
Method	Position ranges (mm)			Angular ranges (degrees)		
	t_x	t_y	t_z	α	β	γ
IC	-2.2, 2.2	-3.0, 1.8	-5.4, 3.6	-18.0, 12.8	-12.4, 13.2	-10.0, 18.2
FC	-2.6, 3.0	-2.8, 3.4	-5.8, 3.8	-10.0, 7.6	-8.4, 7.2	-11.6, 10.8
NMI	-0.6, 1.4	-2.1, 0.8	-5.0, 3.0	-11.6, 2.4	-12.0, 6.8	-11.6, 5.6

Single-Frame Registration Results for Patient A. As in Table II, the Values Denote the Perturbation Ranges That Allowed Successful Registration Over All Patient-A Frames Per Criteria (29) (a) Virtual-to- Virtual Results (b) Virtual-to-Real Results

Table III

(a)						
Method	Position ranges (mm)			Angular ranges (degrees)		
	t_x	t_y	t_z	α	β	γ
IC	-5.0, 8.3	-3.8, 6.5	-8.8, 7.0	-31.5, 32.0	-36.0, 34.0	-56.0, 47.0
FC	-4.3, 4.3	-3.8, 4.3	-4.5, 4.5	-20.0, 26.5	-24.5, 18.5	-43.5, 44.5
NMI	-4.8, 5.5	-3.3, 6.3	-8.3, 8.3	-17.0, 7.0	-18.5, 8.0	-22.5, 14.0

(b)						
Method	Position ranges (mm)			Angular ranges (degrees)		
	t_x	t_y	t_z	α	β	γ
IC	-4.2, 4.8	-3.4, 5.1	-5.6, 5.8	-25.6, 27.8	-33.6, 27.3	-36.2, 18.2
FC	-3.7, 3.9	-3.6, 3.8	-4.7, 5.2	-20.0, 25.1	-26.0, 19.8	-51.1, 44.9
NMI	-4.6, 3.2	-2.1, 4.4	-7.1, 8.6	-12.7, 6.7	-13.8, 8.0	-14.7, 11.1

Table IV

Median Errors for Single-Frame Registration Under Random Initial Pose Perturbations (See Text). Patient A Data Used. Success Rate Indicates the Percentage of Trials That Resulted in Successful Registration Per (29) (a) Virtual-to-Virtual Results (b) Virtual-to-Real Results

(a)					
Method	e_p	e_d	e_r	e_n	success rate
IC	1.6 mm	3.1°	1.7°	1.2 mm	84%
FC	1.7 mm	4.5°	9.7°	1.5 mm	59%
NMI	5.7 mm	8.9°	22.0°	5.7 mm	15%
(b)					
Method	e_p	e_d	e_r	e_n	success rate
IC	2.7 mm	5.4°	8.2°	2.3 mm	70%
FC	2.5 mm	6.3°	10.4°	2.0 mm	52%
NMI	5.4 mm	8.2°	17.1°	5.3 mm	28%

Table V

Virtual-to-Virtual Continuous Registration Results for Two Patient-A Airway Paths. Airway-Path Specifications and Mean Registration Errors Along Each Path are Given

Path	No. of Frames	Path Length (mm)	No. of Airways	e_v	e_d
left lung	225	203	8	1.4 mm	3.1°
right lung	177	148	4	0.7 mm	1.0°

TABLE VI

Virtual-to-Real Continuous Registration Results for Real Bronchoscopic Video Sequences. For Each Sequence, we List the Number of Consecutive Video Frames Successfully Processed Before an Irrevocable Registration Failure Occurred, as Well as Airway-Path Specifications

Sequence	No. of Frames	Path Length (mm)	No. of Airways
patient A, seq. 1	99	30	3
patient A, seq. 2	166	81	4
patient B, seq. 1	345	48	5
patient B, seq. 2	614	55	4
patient C	1,940	52	3
patient D	938	70	3

TABLE VII

Navigation Paths Considered During the Phantom Studies. For Each Traversed Path, we List the Number of Consecutive Video Frames Successfully Processed Before an Irrevocable Registration Failure Occurred. We Also Give the Number and Range of Each Distinct Airway Branch (Generation) Traversed for a Given Path (Trachea = Generation)

Test	No. of Frames	Path Length (mm)	No. (Generation Range) of Airways
Phantom A, path 1, right lung	392	130	8 (3–10)
Phantom A, path 2, right lung	310	106	3 (2–4)
Phantom E, path 1, left lung	274	132	3 (1–3)
Phantom E, path 2, left lung	210	92	4 (2–5)
Phantom E, path 3, right lung	408	103	4 (2–5)

TABLE VIII

Patient Profile for the Mediastinal Nodal-Staging Study. The Study Employed the Integrated Planning and Guidance System

Patient	Age	Gender	Number of Guided Targets
1	54	Male	none
2	62	Female	none
3	59	Male	1 Phantom target
4	50	Male	1 LN + 2 phantom targets
5	39	Female	2 lymph nodes
6	41	Male	1 lymph node, 1 diffuse mass
7	53	Male	1 (Low Patient Oxygen)
8	44	Female	2 lymph nodes
9	77	Male	5 lymph nodes, 1 infiltrate
10	55	Male	2 lymph nodes