

System for Live Virtual-Endoscopic Guidance of Bronchoscopy

J.P. Helferty,¹ A.J. Sherbondy,² A.P. Kiraly,³ and W.E. Higgins⁴

¹Lockheed-Martin Corporation, King of Prussia, PA

²Dept. of Radiology, Stanford University, Stanford, CA

³Siemens Corporate Research Center, Princeton, NJ

⁴Dept. of Electrical Engineering, Penn State University, University Park, PA 16802 USA

Abstract

The standard procedure for diagnosing lung cancer involves 3D computed-tomography (CT) assessment, followed by interventional bronchoscopy. In general, the physician has no link between the CT assessment results and the follow-on bronchoscopy. Thus, the physician essentially performs bronchoscopic biopsy of suspect cancer sites blindly. We have devised a computer-based system that greatly augments the physician's vision during bronchoscopy. The system uses techniques from computer graphics and computer vision to enable detailed 3D CT procedure planning and follow-on image-guided bronchoscopy. The procedure plan is directly linked to the bronchoscope procedure, through a live fusion of the 3D CT data and bronchoscopic video. During a procedure, the physician receives considerable visual feedback on how to maneuver the bronchoscope and where to insert the biopsy needle. We have performed a series of controlled phantom and animal tests, in addition to using the system on a large number of human lung-cancer patients. Results indicate that not only is the variation in skill level between different physicians greatly reduced, but that their accuracy increases.

Keywords—virtual endoscopy, image-guided surgery, 3D imaging, image registration, lung cancer

1. Introduction

Lung cancer is the most common cause of cancer death in the United States [1]. The diagnosis of lung cancer occurs in two stages: (1) 3D computed-tomography (CT) image assessment; and (2) bronchoscopy [2,3]. During 3D CT image assessment, the physician manually “reads” a patient’s 3D CT chest scan to identify and plan bronchoscopic biopsy. This reading is done by either examining a film series of the 3D image data on a view panel or by manually scrolling through the 3D image data on a computer console. Next, during bronchoscopy, the physician attempts to maneuver the bronchoscope through the airways to each preplanned biopsy site. The bronchoscope provides a real-time video stream of the airway interior to assist in this maneuver. Unfortunately, the physician must essentially perform the procedure

blindly, since the target biopsy sites, be they lymph nodes or suspect cancer nodules, are not visible in the local airway video. In addition the physician must make judgments based on the patient’s anatomy depicted in the 3D CT image data; this is difficult as the CT data differs greatly in form from the bronchoscopic video. Thus, physicians vary greatly in their skill level in bronchoscopy and the success rate of bronchoscopic biopsy tends to be very low [2,4].

We describe a computer-based system that improves the accuracy of bronchoscopy and reduces the skill-level variation between different physicians. The system enables detailed 3D CT-based procedure planning and follow-on image-guided bronchoscopy. During both the planning and bronchoscopy stages, the system greatly augments the physician’s vision of the patient’s anatomy. By using computer graphics and other computer-vision methods, far greater use is made of the 3D CT data during procedure planning. During bronchoscopy, the system gives direct image guidance by employing (a) image fusion of the 3D CT data and bronchoscopic video and (b) 3D navigation paths to the preplanned biopsy sites.

Our system has been partly motivated by recent efforts in image-guided surgery [5–7]. These systems add image-based guidance during surgery to improve the procedure success rate and enable difficult procedures. All of these systems require registration of preoperative medical imaging data and the live 3D physical surgical space. To facilitate registration, these systems often employ additional devices, such as fiducial markers attached to the patient’s skin [7], markers attached to the surgical device [6,7], and either an optical or electromagnetic tracking system for measuring the marker positions [6,7]. Our system uses image fusion between the 3D CT image data and bronchoscopic video to perform registration between the preoperative image data and live 3D surgical space, similar to [5]. Hence, we do not require other devices.

Our system also has been partly motivated by the field of virtual endoscopy, which has developed for more exhaustive 3D radiologic image assessment [8]. When applied to the chest, virtual endoscopy is usu-

ally referred to as virtual bronchoscopy (VB) [2–4,8]. In VB, a high-resolution 3D CT chest image serves as a “virtual environment” representing the chest anatomy. Endoluminal (interior) renderings of the airways, generated from computer processing of the CT data, act as views from a “virtual bronchoscope.” Several recent efforts have drawn upon VB to assist bronchoscopy [2–4,9], but these efforts either: (a) did not offer direct image-guided bronchoscopy; (b) only gave information at airway-branch junctions; (c) required far too much computation to be usable during a live procedure; and/or (d) were not tested during live procedures. Notably, though, two of these efforts proposed methods that registered the 3D CT volume (the “*Virtual World*”) to the bronchoscopic video (the “*Real World*”) [4,9] — this enables potentially advantageous fusion of the two image sources without employing a supplemental guidance device. Our system uses fusion of preoperative image data with the 3D physical space.

Sections 2 and 3 of this paper describe our system and specific mathematical details. Section 4 gives results, while Section 5 offers concluding comments.

2. System Overview

We first overview the system. The system is used in two stages, per the standard lung-cancer assessment protocol. Section 3 gives details on many of the data-processing steps. As many choices had to be made during the construction of this system, no individual method can be construed as being “optimal.”

2.1. Stage 1: 3D CT-based Planning

Given a patient’s 3D CT image scan, the airway tree is first segmented using a 3D technique based on region growing and mathematical morphology. Next, using the segmented airway tree as input, the major central axes of the airways are computed using a method combining techniques from 3D skeletonization, branch pruning, and cubic-spline analysis. The techniques for segmentation and central-axes analysis were previously devised in our laboratory and heavily validated on many 3D human CT scans [10].

Next, two sets of triangles representing the interior (endoluminal) and exterior surfaces of the airway tree are generated. These data, necessary for generating 3D renderings of the airway tree, are computed as follows. First, a gray-scale voxel-based mask of the airway-tree surfaces is constructed by combining a simple $5 \times 5 \times 5$ dilation of the airway-tree segmentation with the original 3D gray-scale image data. Next, we apply the standard Marching Cubes algorithm, employed in computer graphics, to this mask to produce the requisite sets of triangles [11].

Finally, the physician interacts with the system’s

computer display to define the target biopsy sites. The physician does this by either manually drawing sites on 2D slice views of the 3D CT data or by employing a semi-automatic image-segmentation method. Once a site is defined, a guidance path is selected by locating the closest precomputed central axis to the site. When this operation is concluded, a guidance plan, consisting of the defined 3D biopsy sites and associated guidance paths, is saved and available for subsequent bronchoscopy.

2.2. Stage 2: Image-Guided Bronchoscopy

During a guided bronchoscopy procedure, our system simultaneously draws upon both the bronchoscope’s video stream and patient’s 3D CT image. In the bronchoscopy laboratory, the bronchoscope’s video feed is interfaced to the computer, giving a live video stream.

During the procedure, the following steps are performed for each preplanned biopsy site. First, the computer display presents the physician with an initial CT rendering along the guidance path. Next, the physician moves the scope “near” the presented site. An automatic registration step is then performed to adjust the virtual CT world to the real video world, bringing the two worlds into registration. When registration is complete, a rendition of the target biopsy site is fused onto the video view and distance information related to the scope’s current position and biopsy-site position is presented. With the aid of the computer, the physician continues down the guidance path to the biopsy site, iterating the steps above. This continues until the physician reaches the biopsy site and performs the biopsy.

The key step during guided bronchoscopy is the registration of the 3D CT to the bronchoscopic video. The bronchoscopic video—the *Real World*—is a live real manifestation of the patient’s chest during the procedure. The 3D CT image—the *Virtual World*—acts as a high-resolution copy of the patient’s chest.

The registration problem can be looked upon as one of matching the viewpoints of two cameras. The first camera—the bronchoscope—gives 2D endoluminal airway video images $I_V(x, y)$ inside the Real World of the human chest. The second camera provides 2D rendered endoluminal airway images $I_{CT}(x, y)$ inside the Virtual World of the CT image. Both cameras provide information, albeit in slightly different forms, on the same physical 3D structure: the interior of the 3D airway tree. The goal of registration is to align the viewpoints of the two cameras so that they are situated at the same point in space and simultaneously give images of the same region.

The registration process is initialized by assuming

that the bronchoscope (Real World camera) is at a fixed viewpoint, giving a fixed reference video image $I_V^F(x, y)$, while the Virtual World camera begins at an initial viewpoint χ_i that is “within a reasonable vicinity” of the bronchoscope’s viewpoint, giving view $I_{CT}^i(x, y)$. During registration, an optimization process searches for the optimal viewpoint χ_o via

$$\chi_o = \arg \left\{ \max_{\chi \in N_{\chi_i}} [S_{NMI} (I_{CT}^X(x, y), I_V^F(x, y))] \right\} \quad (1)$$

to give the virtual image $I_{CT}^{\chi_o}(x, y)$ best matching the fixed video target $I_V^F(x, y)$; in (1) N_{χ_i} represents a search neighborhood about the starting viewpoint χ_i and S_{NMI} represents the normalized mutual information (NMI) between views of the two cameras [12]. Section 3 fully describes the registration problem (1).

3. Mathematical Methods

Section 3.1 describes the shared camera geometry assumed for both the virtual CT world and the real bronchoscopic video world. Section 3.2 discusses image modelling considerations specific to the bronchoscopic video images $I_V(x, y)$, while Section 3.3 describes details related to computing the virtual-world endoluminal views $I_{CT}(x, y)$. Finally, mathematical considerations pertaining to the NMI-based registration problem (1) are given in Section 3.4.

3.1. Camera Geometry

Each data source, $I_V(x, y)$ and $I_{CT}(x, y)$, acts as a camera that provides a 2D image of an observed 3D scene. As discussed below, our system sets up both cameras to abide by the same imaging geometry.

What a camera sees is determined by its viewpoint, specified by the six-parameter quantity $\chi = (X, Y, Z, \alpha, \beta, \gamma)$. (X, Y, Z) represents the camera’s 3D global spatial position in World coordinates, while (α, β, γ) are Euler angles describing the camera orientation about the focal point. A local coordinate system (x, y, z) can be set up about World point (X, Y, Z) . For the local system, the positive z axis is in front of the camera, the positive x axis points to the right, and the positive y axis points up. World point (X, Y, Z) , which is point $(0, 0, 0)$ in local coordinates, coincides with the camera’s focal point; α , β , and γ are the rotation angles about the x , y , and z axes, respectively. The camera’s viewing screen, which captures the resultant 2D image, is perpendicular to the camera’s z axis and is situated a distance f from the focal point, where f is the focal length.

The observed 3D scene is projected onto the camera’s viewing screen through a standard 3D-to-2D perspective projection. For a given observable World point $\mathbf{p} = (X_p, Y_p, Z_p)$, we first transformed it into

the camera’s local coordinate system:

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \end{bmatrix} = \mathbf{R}(\alpha, \beta, \gamma) \begin{bmatrix} X_p - X \\ Y_p - Y \\ Z_p - Z \end{bmatrix} \quad (2)$$

where (X_c, Y_c, Z_c) is the transformed point and $\mathbf{R}(\alpha, \beta, \gamma)$ is the rotation matrix. Finally, the point is converted to a 2D viewing screen location (x, y) through the perspective transformation

$$x = \frac{f X_c}{Z_c}, \quad y = \frac{f Y_c}{Z_c} \quad (3)$$

The viewing screen’s focal length f and physical dimensions determine a camera’s field of view (FOV). To facilitate straightforward registration, we make both the bronchoscope and virtual-world cameras have the same FOV. Thus, if the two cameras are perfectly registered, then pixel (x, y) in bronchoscope image $I_V(x, y)$ and virtual-world image $I_{CT}(x, y)$ arises from the same physical 3D scene point.

To match the FOVs of the two cameras, we do two things. First, prior to bronchoscopy, we calculate the bronchoscopic camera’s focal length f (Section 3.2) and use f for the virtual-world camera’s geometry (Section 3.3). Second, we make the World coordinate system coincide with the 3D CT image’s voxel coordinates. Let the intensity value of voxel (i, j, k) in the 3D CT image be given by $I(i, j, k)$, where i , j , and k are the column, row, and slice indices of the 3D CT image. Then, the World coordinate position of CT voxel (i, j, k) is given by

$$(X_i, Y_j, Z_k) = (i\Delta x, j\Delta y, k\Delta z) \quad (4)$$

where Δx , Δy , and Δz are the sampling intervals.

3.2. Bronchoscopic Image Modeling

A bronchoscope uses a built-in illumination source and a CCD camera to produce a continuous 2D video stream of the observed airway-tree interior. The apertures of these devices are situated at the tip of the bronchoscope. The tip denotes the 3D World position of the bronchoscope inside the airway tree. Most modern CCD-based bronchoscopes produce pseudo-color images. For our work, we need only the luminance (gray-scale) component. Optically, the bronchoscope tip can be modeled as a point light source that coincides with the device’s CCD camera viewpoint [13]. Within this model, the illuminated endoluminal airway surface is Lambertian (diffuse), the image brightness (irradiance) of an illuminated airway surface point $\mathbf{p} = (X_p, Y_p, Z_p)$ is

$$I(\mathbf{p}) = \sigma \frac{L \cos \theta_s}{\pi R^2} \quad (5)$$

where L is the intensity of the bronchoscope’s light source, θ_s is the angle between the light source (same

as the camera’s z axis) and \mathbf{p} ’s surface normal, R is the distance from the light source to \mathbf{p} , and σ is a proportionality factor that takes into account airway-surface albedo and the bronchoscope’s device characteristics. The value $I(\mathbf{p})$ then passes through the bronchoscope camera’s imaging optics, per (2-3), to give the final value $I_V(x, y)$. Bronchoscopic video does tend to have some specular component as well, but this tends to be in small, wet areas and has little impact on the overall scene illumination model.

In reality, a bronchoscope’s camera employs a barrel distortion to give a wide angle (“fish eye”) FOV. This feature gives the physician more detail near the center of the image. Since the barrel distortion literally stretches the observed scene nonlinearly, it has become common to correct for this distortion [14]. For our system, prior to bronchoscopy, we perform a simple off-line calibration that gives a corrective transformation for undoing this barrel distortion. More importantly, this transformation also enables real-time matching of the bronchoscope’s FOV to the virtual-world camera’s FOV [15].

To do this prior off-line computation, the bronchoscope is first mounted in a calibration device at a known distance from a predefined calibration dot pattern, and a bronchoscope image of the dot pattern is captured. Next, a series of calculations are run on the captured (distorted) dot-pattern image. These calculations provide a set of polynomial coefficients that define the distortion-correction transformation.

These calculations also give the focal length f of the bronchoscope camera as follows. Let X_r and X_l denote the horizontal positions of the right-most and left-most dots in the calibration dot pattern, and let x_r and x_l denote analogous quantities for the distortion-corrected image of the pattern. Then, from the perspective equations (2-3),

$$x_r = \frac{f(X_r - X_m)}{Z_m} + x_m \quad x_l = \frac{f(X_l - X_m)}{Z_m} + x_m \quad (6)$$

where x_m is the horizontal coordinate of the viewing-screen center, X_m is an analogous coordinate on the original dot pattern, and Z_m is the known distance of the calibration pattern from the bronchoscope. From (6), the viewing-screen width of this image is $x_r - x_l = \frac{f(X_r - X_l)}{Z_m}$. Thus, the focal length is given by

$$f = \frac{(x_r - x_l)Z_m}{(X_r - X_l)}. \quad (7)$$

All quantities X_r , X_l , X_m , Z_m , x_r , x_l , and x_m are known or easily computed using the calibration calculations. Also, the bronchoscope’s field-of-view angle is readily computed:

$$\theta_{\text{FOV}} = 2 \tan^{-1} \left(\frac{x_r - x_l}{2f} \right). \quad (8)$$

Let x'_r and x'_l denote the horizontal coordinates of the right-most and left-most dots in the distorted (uncorrected) calibration pattern image. The correction-polynomial coefficients are scaled by the factor

$$\frac{x_r - x_l}{x'_r - x'_l} \quad (9)$$

so that the corrected image has the same width as the uncorrected image. Now, the distortion-corrected image will fit into the same window size as the incoming distorted bronchoscopic image. Also, the FOV angle is known.

The bronchoscope’s focal length (7) and FOV angle (8) will be used for the Virtual-World camera, as discussed in Section 3.3. This results in two images sources, I_V and I_{CT} , arising from cameras having matching FOVs.

3.3. Endoluminal Rendering

The 3D CT image defines the Virtual-World representation of the chest. During image-guided bronchoscopy, the Virtual-World camera is maneuvered through the major airways depicted in the 3D CT image. At each viewpoint $\chi = (X, Y, Z, \alpha, \beta, \gamma)$, an endoluminal rendering $I_{CT}^X(x, y)$ is produced. These endoluminal renderings act as simulated endoscopic views of airway-tree interior.

The intensity value $I_{CT}^X(x, y)$ of a particular screen point (x, y) abides by the endoscope Lambertian shading model of (5). For a particular viewpoint χ , a range map and a set of angles θ_s for all needed 3D scene points \mathbf{p} within the required FOV are computed. Calculation of the angles θ_s use the triangle surface normals. The viewing-screen dimensions are determined by the FOV angle θ_{FOV} from (8) and World coordinate-system dimensions from (4). The computation of the viewing-screen image uses the imaging geometry of (2-3) and the focal length (7). We perform all endoluminal view calculations in hardware using standard OpenGL commands.

3.4. Registration Algorithm

The two data sources do not provide physically identical images, but they do have much in common. For the image-guided bronchoscopy problem, Figure 1 shows examples of a video image $I_V^X(x, y)$ and endoluminal rendering $I_{CT}^X(x, y)$ observing the 3D World from the same viewpoint χ . Looking at the two images, it is obvious that they are aligned to view the same 3D structure. As described earlier, both image sources abide by the same imaging geometry and have the same FOV. Also, their intensity characteristics are similar: both sources depict surface-shape information in the form of a depth-shaded Lambertian-surface model [13]. The bronchoscopic video image, however, deviates somewhat



Fig. 1. Matching sample views $I_V(x, y)$ (left) and $I_{CT}(x, y)$ (right) for a typical interior airway location.

from this simple intensity model in that it is able to depict airway-wall mucosal detail and small specular reflections near wet portions of the airway wall. The endoluminal rendering, on the other hand, does not deviate from the model and only depicts surface-shape information, albeit with high quality. Also, a small ambient intensity bias tends to exist between the two images, as a result of rendering options and bronchoscope gain characteristics. The similarities between the two image sources, however, make them well-suited for image registration.

Mutual information, which arises in information theory to measure the statistical dependence between two random variables, is commonly used for registering images from different modalities [5, 12, 16, 17]. It can be measured with the Kullback-Leibler metric [17], which for our problem is given by

$$S_{MI}(I_V, I_{CT}) = \sum_{k=0}^{M-1} \sum_{l=0}^{M-1} p_{V,CT}(k, l) \log \frac{p_{V,CT}(k, l)}{p_V(k) \cdot p_{CT}(l)} \quad (10)$$

where images I_V and I_{CT} are the two “random variables” being compared, $p_V(k)$ and $p_{CT}(l)$ are the respective marginal probability density functions of the images (normalized histograms), $p_{V,CT}(k, l)$ is the joint density function between the two images, and $M = 256$ is the number of gray-levels used. S_{MI} can also be written in terms of entropy:

$$S_{MI}(I_V, I_{CT}) = h(V) + h(CT) - h(V, CT), \quad (11)$$

where $h(V)$ and $h(CT)$ are the marginal image entropies and $h(V, CT)$ is the joint entropy between the two images.

Studholme *et al.*, however, performed a detailed study demonstrating that the basic S_{MI} measure can fail to properly align two images if the amount of image overlap is large [12]. Also, in bland overlapping image regions, where $h(CT) \approx h(V) \approx h(V, CT)$, S_{MI} is sensitive to the lack of statistical information contributed by these overlapping regions. This reduces the measure’s ability to recover from larger initial misalignments and to register images that have much overlap. For these reasons, we use the idea of normalized mutual information (NMI) by Studholme

et al. [12], which involves normalizing S_{MI} in (11) by $h(V, CT)$ and removing the constant term:

$$S_{NMI}(I_V, I_{CT}) = \frac{h(V) + h(CT)}{h(V, CT)} \quad (12)$$

For S_{NMI} , any increase in marginal entropies is counterbalanced by a change in joint entropy, making the measure less dependent on the amount of overlap [12].

Bricault *et al.* have pointed out for the endoluminal airway images that [4]: (1) the significant information resides primarily near airway bifurcations and the corresponding dark “holes” leading to upcoming airways and near large (darker) deviations in the airway walls; and (2) the brighter bland wall regions tend to have little useful information. Also, the specular reflections that appear in small regions of the video images correspond to saturated intensity points and do not appear in the endoluminal renderings. Drawing upon these observations, we modify the entropy calculations by varying the weight assigned to darker and brighter image points. Noting that $p_V(k) = \sum_{l=0}^{M-1} p_{V,CT}(k, l)$ and $p_{CT}(l) = \sum_{k=0}^{M-1} p_{V,CT}(k, l)$, we can write

$$\begin{aligned} h(V) &= - \sum_{k=0}^{M-1} \sum_{l=0}^{M-1} p_{V,CT}(k, l) \log p_V(k) \\ h(CT) &= - \sum_{k=0}^{M-1} \sum_{l=0}^{M-1} p_{V,CT}(k, l) \log p_{CT}(l) \\ h(V, CT) &= - \sum_{k=0}^{M-1} \sum_{l=0}^{M-1} p_{V,CT}(k, l) \log p_{V,CT}(k, l). \end{aligned}$$

These are modified by adding weighting factors:

$$\begin{aligned} h(V) &= - \frac{1}{M^2} \sum_{k=0}^{M-1} \sum_{l=0}^{M-1} w_k w_l p_{V,CT}(k, l) \log p_V(k) \\ h(CT) &= - \frac{1}{M^2} \sum_{k=0}^{M-1} \sum_{l=0}^{M-1} w_k w_l p_{V,CT}(k, l) \log p_{CT}(l) \\ h(V, CT) &= - \frac{1}{M^2} \sum_{k=0}^{M-1} \sum_{l=0}^{M-1} w_k w_l p_{V,CT}(k, l) \log p_{V,CT}(k, l) \end{aligned} \quad (13)$$

where the weights w_k are given by

$$w_k = M - 1 - k, \quad k = 0, 1, \dots, M - 1 \quad (14)$$

(The w_l are also given by (14) with l replacing k .) While other weighting schemes are possible, this scheme emphasizes darker pixels while attenuating brighter pixels, where saturated points at gray-level $M - 1$ are deleted. Thus, greater emphasis is placed on the important darker image structures.

For our circumstance, S_{NMI} is maximized when the joint density $p_{V,CT}(k, l)$ is approximately a diagonal matrix. This behavior can be explained

as follows. If the two image sources are perfectly aligned and have identical intensity characteristics, then $\forall k, p_{V,CT}(k, k) = p_V(k) = p_{CT}(k)$. But, as stated earlier, the two image sources differ somewhat in intensity characteristics. If the images become properly aligned, then $P_{V,CT}(\cdot, \cdot)$ will still be concentrated along a diagonal in the $k-l$ space, with some deviation about the diagonal to account for small local deviations in source differences and with a possible shift to account for ambient intensity bias.

Finally, to solve the registration problem (1), the optimization starts with initial viewpoint χ_i for the endoluminal renderings and with fixed target video frame $I_V^F(x, y)$. During optimization, viewpoint $\chi = \{X, Y, Z, \alpha, \beta, \}$ is varied in a neighborhood N_{χ_i} about χ_i . An endoluminal rendering $I_{CT}^X(x, y)$ is computed for each candidate viewpoint χ and compared to $I_V^F(x, y)$ through the S_{NMI} measure (12), weighted entropies (13), and weights (14). This is a six-parameter optimization problem, per the six parameters constituting χ . χ is varied until an optimal viewpoint χ_o is found that maximizes S_{NMI} . Three standard optimization methods were investigated: (1) a steepest-ascent, (2) Nelder-Mead simplex, and (3) simulated annealing. In our experiments, all methods converged well in the large majority of circumstances. The simplex method and simulated annealing, however, were least susceptible to local minima. We have generally used the simplex method for live human procedures, as it is less computationally intensive than simulated annealing.

4. Results

This section presents a series of phantom, animal, and human results. Our system uses a standard Dell Precision 620 workstation PC (dual-933MHz Pentium-III, 2GB RAM, Windows 2000). A Matrox Meteor-II frame grabber board is used for real-time video capture, while the GeForce4 Ti4600, which has an onboard 128MB video memory, is used for the video card. This inexpensive computer set-up provides real-time computation of high-quality endoluminal renderings and real-time presentation of the bronchoscopic video. Figure 2 depicts the system in the surgical suite. The physician observes both the computer display and standard fail-safe bronchoscope video monitor during the procedure.

We first performed a controlled study, involving no motion. Six physicians, having a wide range of experience, performed bronchoscopic ‘‘biopsy’’ on a rubber model of the airway tree augmented with five 1.4mm platinum beads (the biopsy sites). The phantom was made of rigid rubber and derived from a human model. A CT scan was done of the phan-



Fig. 2. System usage during a live procedure.

tom. The system was used to plan biopsy to the target sites. Also, a standard CT film was made. The physicians then performed bronchoscopy of the five sites with the film only (standard practice) and with the guidance system. The results are as follows:

Physician	standard (mm)	guided (mm)
1	5.80	1.38
2	2.73	1.33
3	4.00	1.49
4	8.87	1.60
5	8.62	2.45
6	3.19	1.24

The average biopsy error over the six physicians for the standard film-only case was 5.53mm, with a standard deviation of 4.36mm. Using the guidance system, this average error decreased dramatically to 1.58mm, with a standard deviation of 1.57mm. Notably, physician performance almost appeared to become independent of experience, and they all performed nearly the same! Considering that typical target sites, such as lymph nodes and suspect cancer lesions, typically have a diameter > 1 cm, this is an excellent result [2]. Note that typical biopsy needles have length 20mm and diameter 0.5-1.0mm.

We next performed a five animal studies to determine the system’s effectiveness during a live procedure. We first performed a 3D CT scan of the animal and planned a series of ‘‘virtual’’ biopsy sites. We then used the guidance system to guide the physician to each of the biopsy sites. Upon reaching a site, the physician would deposit a metallic dart. Figure 3 depicts a sample computer display during a test. When all sites had been ‘‘biopsied,’’ we then rescanned the animal and compared the before and after CT scans—Figure 4 depicts sample comparison results. The average distance of each dart from the center of the preplanned site was 2.3mm.

Figures 5-6 show results for human lung-cancer patient h005. Figure 5 depicts results of the 3D CT Assessment stage, while Figure 6 shows a computer screen shot during a procedure. Extensive display

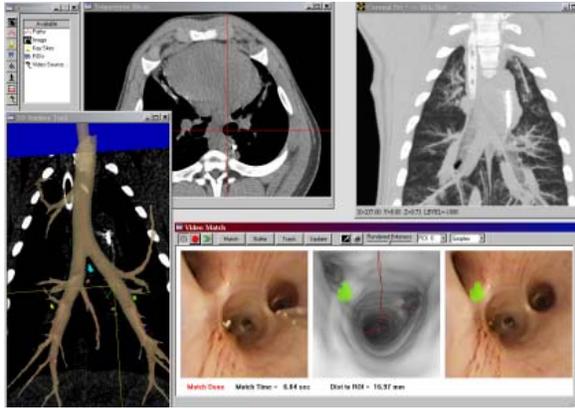


Fig. 3. Computer display during an animal test. Upper left view is a transverse slice, upper right view is a coronal front-to-back thin-slab rendering, lower right is a 3D surface rendering of the airway tree, central axes, and planned biopsy sites (small blobs in color). The lower right Video Match view shows the results during live registration: (a) left view is live bronchoscopic video inside the airways; (b) center view is registered CT-based endoluminal rendering; (c) right view shows the green CT-based biopsy site fused onto the registered video view. Note that the green site is beyond the airway walls and not visible in the real video. By doing the image fusion, we provide the physician with an augmented vision of the situation.

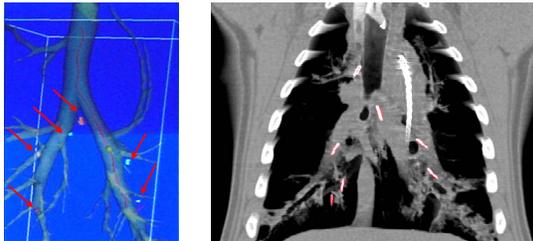


Fig. 4. Results of animal test. Left view: 3D surface rendering showing segmented airway tree, central axes (thin lines), and six planned biopsy sites (pointed to by arrows). Right view: coronal thin-slab rendering showing metallic darts (appear as bright whites) deposited at preplanned biopsy sites.

capability is provided to the physician by the various graphical tools to give a much fuller vision of the procedural circumstances.

Finally, Figure 7 focuses on the Video Match tool for human lung-cancer case DC. In the figure, the physician has introduced a biopsy needle into one of the bronchoscope’s working channels. The needle is shown as the physician is about to make a biopsy of a target site. Since the needle is bright, it has minimal impact on the registration (the registration emphasizes dark areas).

5. Discussion

A vital assumption made is that the “real world” 3D space captured by the bronchoscopic video is in synchrony with the “virtual world” 3D space depicted by the CT renderings. This implies that the physi-

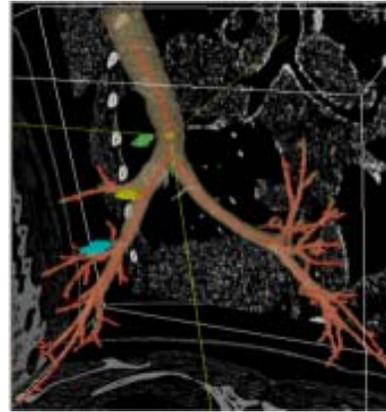


Fig. 5. Stage-1 3D CT Assessment for human case h005. 3D surface rendering depicts segmented airway tree, extracted central axes (red lines), and three preplanned biopsy sites (green, yellow, and blue blobs). Scan done on 4-detector Philips Mx8000 MDCT scanner. $\Delta x = \Delta y = 0.59\text{mm}$, $\Delta z = 0.60\text{mm}$; 479 slices, each 512×512 , constitute the

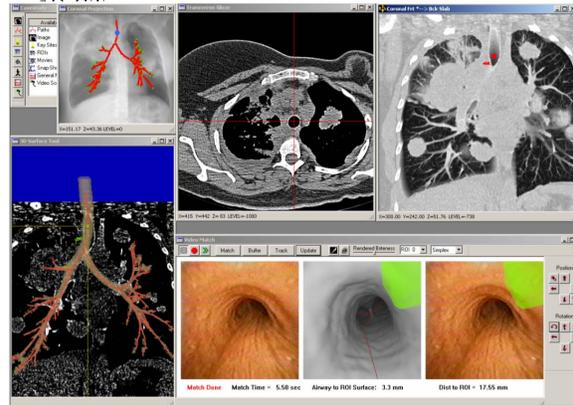


Fig. 6. System view during Stage-2 Image-Guided Bronchoscopy for human case h005. Upper right view shows weighted-sum projection of 3D CT image data, plus the projected red central axes—current chest location is marked with a blue ball. Upper center and right views show a transverse 2D CT slice and coronal front-to-back thin-slab rendering—cross-hairs and red ball indicate current 3D chest location. Lower left 3D surface rendering depicts airway tree, central axes, and current green biopsy site—a needle indicates current 3D position. Lower right Video Match view shows the following: live video at current 3D location, registered CT-based endoluminal rendering with green biopsy site and red-line guidance path (note that the green site is *outside* the airway), video view with CT-based biopsy site fused onto it; distance information provided by this tool states how far the bronchoscope currently is from the biopsy site.

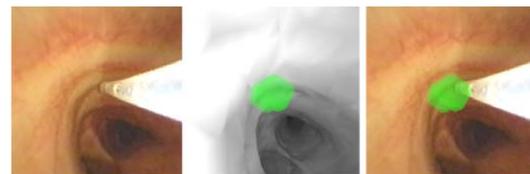


Fig. 7. Blow up of Video Match tool view during actual biopsy for human case DC. Note the needle in the field of view piercing the virtual biopsy site in the far-right bronchoscopic video frame.

cian “cooperates” when he moves the scope close to a biopsy site. Through a detailed study, we have found that the registration technique behaves robustly over a wide range of translations and rotations, giving the physician much leeway in the level of cooperation.

We have also found that the time difference between when the CT scan is done and when the procedure is performed also does not have an impact. In addition, the difference in inspiration level (how full the lungs are with air) between when the CT scan is done and when bronchoscopy is performed also does not have a significant effect. This appears to be true, because the airways are relatively rigid and bony, and their local relative shape and size do not change during inspiration.

In addition to the phantom and animal studies, we have successfully used the system on nearly 40 human lung-cancer patients to date. The system increases the physician’s vision of the procedure circumstances, greatly eases decision making, and appears to increase biopsy success rate. Notably, the system appears to nullify the skill-level difference between different physicians, while also improving accuracy.

The registration procedure is only run at discretely selected sites along a path toward a target biopsy site. For our computer set-up, a registration takes on the order of 5-10 sec. The bulk of this time is taken up by the large number of endoluminal renderings that must be computed during a typical optimization. More efficient implementation can speed up this procedure.

Acknowledgments

This work was partially supported by NIH-NCI grants # R01-CA074325 and R44-CA091534. From Penn State, we would like to thank Janice Turlington, Allen Austin, David Zhang, Dirk Padfield, James Ross, Tao Yang, and Shu-Yen Wan, who contributed to the system and participated in some of the experiments. From the University of Iowa, we would like to thank Eric Hoffman, Geoffrey McLennan, Scott Ferguson, Karl Thomas, Alan Ross, Janice Cook-Granroth, Angela Delsing, Osama Saba, Deokiee Chon, Jered Sieren, and Curt Wolf, who participated in some of the experiments.

REFERENCES

- [1] R. Greenlee, M. Harmon-Hill, T. Murray, and M. Thun, “Cancer statistics, 2001,” *CA Cancer J. Clin.*, vol. 51, no. 1, pp. 15–36, Jan.-Feb. 2001.
- [2] H. P. McAdams, P. C. Goodman, and P. Kussin, “Virtual bronchoscopy for directing transbronchial needle aspiration of hilar and mediastinal lymph nodes,” *Am. J. Roentgen.*, pp. 1381–1364, May 1998.
- [3] K. Hopper, T. Lucas, K. Gleeson, J. Stauffer, R. Bascom, D. Mauger, and R. Mahraj, “Transbronchial biopsy with virtual CT bronchoscopy and nodal highlighting,” *Radiology*, vol. 221, no. 2, pp. 531–536, Nov. 2001.
- [4] I. Bricault, G. Ferretti, and P. Cinquin, “Registration of real and CT-derived virtual bronchoscopic images to assist transbronchial biopsy,” *IEEE Transactions on Medical Imaging*, vol. 17, no. 5, pp. 703–714, Oct. 1998.
- [5] W. E. Grimson, G. J. Ettinger, S. J. White, T. Lozano-Perez, W. E. Wells III, and R. Kikinis, “An automatic registration method for frameless stereotaxy, image guided surgery, and enhanced reality visualization,” *IEEE Transactions on Medical Imaging*, vol. 15, no. 2, pp. 129–140, Apr. 1996.
- [6] Y. Sato, M. Nakamoto, Y. Tamaki, T. Sasama, I. Sakita, Y. Nakajima, M. Monden, and S. Tamura, “Image guidance of breast cancer surgery using 3-D ultrasound images and augmented reality visualization,” *IEEE Trans. on Medical Imaging*, vol. 17, no. 5, pp. 681–693, Oct. 1998.
- [7] J. D. Stefansic, A. J. Herline, Y. Shyr, W. C. Chapman, J. M. Fitzpatrick, B. M. Dawant, and R. L. Galloway Jr., “Registration of physical space to laparoscopic image space for use in minimally invasive hepatic surgery,” *IEEE Transactions on Medical Imaging*, vol. 19, no. 10, pp. 1012–1023, Oct. 2000.
- [8] D. J. Vining, K. Liu, R. H. Choplin, and E. F. Haponik, “Virtual bronchoscopy: relationships of virtual reality endobronchial simulations to actual bronchoscopic findings,” *Chest*, vol. 109, no. 2, pp. 549–553, Feb. 1996.
- [9] K. Mori, D. Deguchi, J. Sugiyama, Y. Suenaga, J. Toriwaki, C.R. Maurer, H. Takabatake, and H. Natori, “Tracking of bronchoscope using epipolar geometry analysis and intensity-based image registration of real and virtual endoscopic images,” *Medical Image Analysis*, vol. 6, pp. 321–336, 2002.
- [10] A. P. Kiraly, E. A. Hoffman, G. McLennan, W. E. Higgins, and J. M. Reinhardt, “3D human airway segmentation methods for virtual bronchoscopy,” *Academic Radiology*, vol. 9, no. 10, pp. 1153–1168, Oct. 2002.
- [11] W. Schroeder, K. Martin, and B. Lorensen, *The Visualization Toolkit: An Object-Oriented Approach To 3D Graphics*, Prentice Hall, Upper Saddle River, N.J., 1997.
- [12] C. Studholme, D. L. G. Hill, and D. J. Hawkes, “An overlap invariant entropy measure of 3D medical image alignment,” *Pattern Recognition*, vol. 32, no. 1, pp. 71–86, Jan. 1999.
- [13] T. Okatani and K. Deguchi, “Shape reconstruction from an endoscope image by shape from shading technique for a point light source at the projection center,” *Computer Vision and Image Understanding*, vol. 66, no. 2, pp. 119–131, May 1997.
- [14] K.V. Asari, S. Kumar, and D. Radhakrishnan, “A new approach for nonlinear distortion correction in endoscopic images based on least squares estimation,” *IEEE Trans. Med. Imaging*, vol. 18, no. 4, pp. 345–354, April 1999.
- [15] J.P. Helferty, C. Zhang, G. McLennan, and W.E. Higgins, “Videoscopic distortion correction and its application to virtual guidance of endoscopy,” *IEEE Trans. Med. Imaging*, vol. 20, no. 7, pp. 605–617, July 2001.
- [16] P. Viola and W. M. Wells III, “Alignment by maximization of mutual information,” *Int. Journal of Comp. Vis.*, vol. 24, no. 2, pp. 137–154, Feb. 1997.
- [17] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens, “Multimodality image registration by maximization of mutual information,” *IEEE Transactions on Medical Imaging*, vol. 16, no. 2, pp. 187–198, Apr. 1997.